



HAL
open science

The power of leadership in changing social norms in heterogeneous societies

Fabio Galeotti, Jona Krutaj, Marie Claire Villeval

► **To cite this version:**

Fabio Galeotti, Jona Krutaj, Marie Claire Villeval. The power of leadership in changing social norms in heterogeneous societies. Proceedings of the National Academy of Sciences of the United States of America, 2026, 123 (13), pp.e2526916123. <10.1073/pnas.2526916123>. <hal-05595719>

HAL Id: hal-05595719

<https://hal.science/hal-05595719v1>

Submitted on 17 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



The power of leadership in changing social norms in heterogeneous societies

Fabio Galeotti^{a,1} , Jona Krutaj^{b,1} , and Marie Claire Villeval^{a,1,2}

Affiliations are included on p. 9.

Edited by Rick K. Wilson, Rice University, Houston, TX; received September 23, 2025; accepted January 29, 2026 by Editorial Board Member Margaret Levi

Abandoning detrimental social norms is complex due to the strong pressure to conform. We examine how leaders can guide norm change in heterogeneous societies where individual preferences evolve at different rates. Inspired by the model and experimental design of [J. Andreoni, N. Nikiforakis, and S. Siegenthaler, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014893118 (2021)], we conduct a large-scale laboratory experiment in which we manipulate the speed at which preferences change within a society and introduce leaders with different, evolving preferences. Without leaders, a minority of citizens with rapidly changing preferences cannot overturn an existing norm in a society where most individuals have slow-changing preferences. When fast-changing citizens form the majority, norm change occurs in most groups, but at high welfare costs. In contrast, exogenously selected leaders are highly effective at coordinating expectations and shifting heterogeneous societies toward a more efficient norm—at lower welfare costs and regardless of the underlying distribution of preference evolution across individuals. However, the timing of norm change depends on whether leaders prioritize their preferences (autocratic leadership) or those of the majority (democratic leadership). A follow-up experiment shows that peer-to-peer communication encourages leaders to adopt a more democratic leadership style. These results highlight the pivotal role of leadership in driving norm change and the importance of public voice in shaping leaders' behavior.

social norm | leadership | coordination | tipping threshold | experiment

Social norms, defined as informal rules of behavior that individuals follow when they expect others in their reference group also to follow, are fundamental for the harmony and the development of societies (1). Although norms generally promote social welfare, some become detrimental over time, sustaining behaviors that impose high costs on society, such as corruption or consumption habits that contribute to climate change (2, 3). In such cases, even if a norm change would be socially beneficial, societies often continue to uphold such norms due to the belief that others continue to follow the norm and that deviating from expected behavior will therefore result in social sanctions—that is, costs, material or psychological, imposed by other members of society on those who violate a norm (4–7).

Spontaneous normative changes are possible, but only when enough individuals abandon the norm, such that incentives reverse across the population, propelling rapid change toward an alternative state (8–17). Andreoni et al. (18) modeled this tipping point as depending on the trade-off between the benefits from change and the costs of deviating from the norm (18). In their setting, individuals' preferences are homogeneous across the population and evolve exogenously over time at the same rate, increasing incentives to adopt an alternative norm. This evolution reflects, for example, learning about superior social alternatives or changes in population composition due to migration or generational turnover. Even if, over time, preferences for adopting an alternative norm become predominant, norm change can occur only if social expectations—that is, individuals' beliefs about how others will act and think one ought to behave—also shift.

Building on the model and experimental design of Andreoni et al. (18), we depart from their assumption of homogeneous individuals to examine settings in which individuals differ in how their preferences evolve over time—an environment that more closely reflects the diversity of real societies, and where coordination on a new norm may be more complex or costly to achieve. This allows us to study how the introduction of a leader, whose preferences can differ from those of the majority as preferences evolve, affects the likelihood of coordinated norm change.

Significance

Social norms, though essential for social order, often persist even when no longer beneficial. In a large-scale laboratory experiment, we investigate whether leaders can promote beneficial change in societies with different rates of preference adjustment. Without leaders, in societies where majority preferences evolve slowly, a more dynamic minority favoring change is insufficient to shift norms. In societies where majority preferences evolve rapidly, norms may shift, but at high welfare costs. Introducing leaders helps coordinate expectations and overturn established norms, even when leaders' preferences temporarily diverge from those of the majority. However, welfare outcomes depend on whether leaders align recommendations with their preferences or those of the majority. Citizen communication increases the weight leaders place on majority preferences in their recommendations.

Author contributions: F.G., J.K., and M.C.V. designed research; J.K. performed research; J.K. analyzed data; and F.G., J.K., and M.C.V. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. R.K.W. is a guest editor invited by the Editorial Board.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹F.G., J.K., and M.C.V. contributed equally to this work.

²To whom correspondence may be addressed. Email: villeval@gate.cnrs.fr.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2526916123/-/DCSupplemental>.

Published March 25, 2026.

In our setting, some individuals have preferences that evolve very slowly, capturing resistance to behavioral change. In contrast, others have preferences that evolve more rapidly, reflecting greater responsiveness to new social alternatives. This variation alone, without yet introducing leaders, raises several important questions that we address in this paper: When most individuals' preferences change very slowly, can a small group of individuals with fast-evolving preferences instigate norm change, or do such societies remain stuck in old, detrimental norms? Conversely, when most individuals' preferences evolve rapidly, can a minority with slow-evolving preferences block normative shifts?

While a society's composition may allow for instigators of change to emerge, coordinating social expectations may necessitate external triggers, such as laws, court decisions, mass communication technologies (19–24) or leaders. Given their prominent role, leaders possess the potential to either promote or resist normative shifts (25–28), particularly in societies composed of groups with different responsiveness to change. Such heterogeneity may also generate temporary conflicts between leaders' preferences and those of the majority, potentially leading to backlash. Illustrative examples of leaders who successfully initiated shifts toward more cooperative and inclusive norms in diverse, divided societies include A. Kemal in Turkey, M. Gandhi in India, or N. Mandela in South Africa. In other instances, such efforts failed due to resistance from the majority, as with B. Obama's attempts to regulate gun ownership or A. Lincoln struggles to abolish slavery. Conversely, leaders can also block change through repression or persuasion, until a majority eventually overcomes their influence (29, 30). Leaders can influence individual behavior through shifting beliefs about how others will act (31–36) or altering normative beliefs about appropriate behavior (37–39). In static coordination settings, they can serve as effective coordination devices, but only when their actions or types are credible—for example, when leaders' incentives are aligned with those of followers (40).

By investigating the effectiveness of leaders in fostering or inhibiting norm change in societies with heterogeneous preference evolution—where differences in responsiveness to change can exacerbate coordination problems and create conflict between leaders' preferences and those of the majority—we open up several important questions: Will leaders defend the majority view when their own preferences are closer to those of a minority, that is, will they adopt a “democratic” leadership style? Or will they prioritize their own preferences over those of the majority and adopt an “autocratic” style, potentially failing to induce change if followers do not respond? Is democratic leadership more effective than autocratic leadership in coordinating social expectations?

A major challenge in assessing the influence of leadership on norm change in real-world contexts is the endogeneity of leader selection, which makes it difficult to distinguish whether the support for a norm is attributed to the influence of a leader or if the leader's communication reflects the collective view, for example, to increase their chances of re-election. For these reasons, we conducted a laboratory experiment where we i) simulated the emergence of norms by inducing incentives that mirror those found in the real world, ii) varied the distribution of preference adjustment, creating societies in which most individuals' preferences evolve relatively quickly or relatively slowly, and iii) exogenously introduced leaders whose preferences initially align with those of the rest of the society, but evolve at a different pace than the majority. By exogenously selecting a

leader and manipulating the pace of preference change, we can isolate the causal effect of leadership on shifting existing norms.

Testing the Impact of Leaders on the Dynamics of Norms

Subjects were assigned to societies of either 20 members (baseline conditions with no leader) or 21 members (leader treatments). In the leader treatments, one subject was randomly designated as the leader, while the remaining 20 participants assumed the role of citizens.

The citizens participated in a dynamic game introduced by Andreoni et al. (18) for 31 periods. In each period, they were randomly paired with another subject and were asked to choose between two options: Blue or Green. Preferences over these two options were induced by assigning a higher payoff of $v_H = 30$ Experimental Currency Units (ECU) to one option and a lower payoff of $v_L = 20$ Experimental Currency Units to the other. To establish an initial social norm, all citizens were initially induced to prefer Blue. In the early periods, if citizens coordinated on Blue, they received a higher payoff than if they coordinated on Green: (BB = v_H, v_H) vs. (GG = v_L, v_L). Failure to coordinate with one's partner triggered a penalty equal to a fixed rate ($P = 4$) multiplied by the number of citizens choosing the other color. For instance, if an individual chose Blue, they incurred a penalty of $p \times g_t$, where g_t corresponds to the proportion of individuals choosing Green in that period. Conversely, a citizen choosing Green incurred a penalty of $p \times (1 - g_t)$. This payoff structure induced a strong pressure to conform to the established norm by imposing higher costs on early deviators (SI Appendix, section A1 and Table S1).

Individual preferences over one of the two options (Blue/Green) shifted over time, generating impetus for norm change. From period 2 onward, in each period, there was an individual cumulative probability that a subject's preference would permanently switch from Blue to Green. This means that, from a certain period onward, citizens received a higher payoff if they coordinated with the others on Green as opposed to Blue: (BB = v_L, v_L) vs. (GG = v_H, v_H). The individual probabilities were publicly known, allowing citizens to anticipate that, over time, everyone would eventually prefer Green, and thus, abandoning the Blue norm in favor of the Green norm would be individually and socially beneficial. However, the emergence of Green as the new norm is not straightforward, as early adopters of Green incur higher costs, and the historical adherence to the now-inefficient Blue norm biases individual expectations toward Blue. In this context, introducing leaders can provide the necessary trigger to alter these expectations and induce social change.

The experimental feedback for citizens included the choice of their paired partner in the previous period, as well as their own earnings. However, they did not observe the preferences of the other citizens in their society or learn how many people had changed their preferences during a given period.

The leader was chosen randomly from the 21 members of the society. Unlike the citizens, they could observe the preferences of all members of the society at the beginning of each period. The leader did not participate in the dynamic coordination game themselves, yet held a prominent role by conveying a visible message to all citizens (41–45). At the start of each period, after observing their own preference, the preferences of all other citizens, and the latter's actions in the previous period, the leader could send one of two messages to the citizens: “I advise you to

choose Blue” or “I advise you to choose Green.” This form of communication was chosen to ensure that the leader’s message was clear and actionable.

The leader’s recommendation acts as a public coordination signal, and, in this respect, is conceptually related to the Preference Poll treatment in Andreoni et al. (18) in which individuals publicly revealed their preferred alternative through a one-time poll. A key difference is that, in our design, the signals are repeatedly issued by a single, identifiable individual whose preferences may temporarily diverge from those of the majority and whose incentives are structurally distinct. This allows us to study how leadership, rather than aggregate preference revelation, modifies an established social norm.

Specifically, the leader was incentivized to convince the citizens to coordinate on the leader’s preferred norm. The leader received a higher payoff if all citizens chose the leader’s preferred norm than if all chose the alternative norm (900 vs. 400 Experimental Currency Units). Additionally, the leader was induced to dislike normative conflict: It was more important for the leader to have all citizens coordinating on any norm, rather than having no norm (a fully divided society with half of the citizens choosing Blue and the other half choosing Green). This setup reflects real-world settings where a leader may predicate an action that deviates from their own preferences to avoid harmful miscoordination. For example, B. Obama and D. Cheney withheld their early then-unpopular support for same-sex marriage until broader societal acceptance made their stance more publicly palatable.

To incorporate this feature, the leader faced penalties for each citizen who deviated from the prevalent norm in the society, with the penalty varying depending on whether the deviant supported the leader’s preferred norm or not. The fixed penalty for the leader was low ($f_L = 20$ ECU) when the deviators chose the norm the leader preferred, and high ($f_L = 90$) when they chose the norm the leader did not prefer. This structure resulted in a V-shaped payoff function for the leader (SI Appendix, Fig. S2).

This approach differs from the Incentive for Instigators condition in Andreoni et al. (18), where explicit rewards for instigators of change were introduced to test whether incentivizing early movers could overcome coordination failures, but without any additional information or communication channel. In contrast, our leaders were selected randomly, occupied a specific institutional role as informational intermediaries, and had a distinct payoff function, rather than merely receiving larger incentives from change than other group members.

Inducing the Change in Preferences

To investigate the impact of leadership across different heterogeneous societies, we varied the type of society (“Slow-evolving” vs. “Fast-evolving”), and the presence and type of the leader (“Slow-evolving” vs. “Fast-evolving”) by manipulating the pace at which preferences changed.

This resulted in four between-subject treatments, as illustrated in Table 1: two baseline treatments with either Slow- or Fast-evolving preferences in the society and no leader, and two additional treatments with the same types of societies interacting with a leader. In the latter cases, we considered leaders whose preferences evolved at a speed opposite to that of the majority in society. This was done to examine the emergence of “autocratic” vs. “democratic” leadership styles and their respective effects on norm change, particularly in situations where coordination problems and conflicts of interest are most salient. Specifically, it was necessary to create a time interval during which the preferences

of the leader and those of the majority were no longer aligned. This divergence was achieved by assigning Fast-evolving leaders to Slow-evolving societies and Slow-evolving leaders to Fast-evolving societies. We denote the former as the Slow-evolving Society-Fast-evolving Leader (SS-FL) treatment and the latter as the Fast-evolving Society-Slow-evolving Leader (FS-SL) treatment. For each treatment, we collected 10 independent societies.

To induce slow- or fast-evolving preferences, we manipulated the probability at which individual preferences evolved over time.

Societies with Slow-Evolving or with Fast-Evolving Preferences.

All citizens were induced to prefer Blue at the beginning of the game. In Slow-evolving societies, 80% of citizens changed their preferences from Blue to Green, with a probability that increased cumulatively only by 0.6 percentage points in each subsequent period, starting from period 2. The remaining 20% experienced a cumulative increase of 6 percentage points per period, representing a Fast-evolving minority. In Fast-evolving societies, the probability of 80% of citizens changing preferences increased cumulatively by 6 percentage points in each period. The probability of the 20% remaining citizens increased cumulatively by 0.6 percentage points per period, representing a Slow-evolving minority. This allows us to examine whether such minorities can become obstacles to the normative progress hoped for by the majority, or, conversely, act as trendsetters capable of initiating a norm change. All subjects knew the probabilities with which their own preference and others’ preferences changed. They also knew that the preference change was permanent: Once an individual started preferring Green, they continued to prefer Green until the end of the game.

According to the selected distributions (SI Appendix, section A2 and Fig. S3), the majority of citizens preferred Green from periods 15 and 7 in Slow-evolving and Fast-evolving societies, respectively. Thus, for the majority, it became beneficial to coordinate on the Green norm from period 15 in Slow-evolving societies and from period 7 in Fast-evolving societies. By the end of the game, nearly everyone was induced to prefer Green in both societies.

Slow-Evolving and Fast-Evolving Leaders. At the beginning of the game, the leader was induced to prefer Blue like the other members of the society, thus being representative of the society. We manipulated whether the leader had slow-changing preferences (“Slow-evolving”) or rapidly changing preferences (“Fast-evolving”). The leader’s probability of changing preferences increased cumulatively by only 0.6 percentage points in each period in the former case, and by 6 percentage points in each period in the latter case. We predetermined the period in which the leader’s preference changed by taking the average of 1,000 draws following the respective probabilities. This ensured that all leaders in a given condition changed preference at the same time: in period 17 for all Slow-evolving leaders and period 6 for

Table 1. Treatments

Treatment	Society type	Leader type	Number of societies (independent observations)
SS-Baseline	Slow-evolving	No leader	10
FS-Baseline	Fast-evolving	No leader	10
SS-FL	Slow-evolving	Fast-evolving	10
FS-SL	Fast-evolving	Slow-evolving	10

all Fast-evolving leaders. Citizens were informed about the type of leader in their society, but not about the period in which the leader's preferences changed.

Measuring Autocratic and Democratic Leadership Styles

In the SS-FL treatment, the leader began to prefer Green in period 6, whereas the majority's preference shifted toward Green from period 15 onward. A leader with an "autocratic" style should advocate for the adoption of the new norm as soon as their own preference shifts, regardless of the majority's inclination toward Blue. In contrast, a leader with a "democratic" style should wait until at least period 15 when the majority aligns with Green, even if the leader's own preference has already transitioned several periods earlier. Conversely, in the FS-SL treatment, the leader only started preferring Green in period 17, whereas the majority had already begun preferring Green from period 7. Here, an autocratic-style leader should condition their recommendation of Green on the shift in their individual preference, that is, from period 17 onward. In contrast, a democratic-style leader should start recommending Green as soon as the preference of the majority switches to Green, that is, several periods before the leader's individual preference changes to Green. Any deviation from these strategies within the defined time frame would characterize a compromising strategy and the extent to which the leader would be willing to make concessions. Fig. 1 provides a visual representation of the democratic and autocratic leadership styles.

Results

Fig. 2 displays the evolution over time of the citizens' preferences and their choice of Green for each society in the baseline treatments. In all societies, Blue emerged as the initial norm, with 97.3% of citizens choosing Blue during the first six periods of the game, when the majority preferred Blue.

Norm Change Without a Leader. In Slow-evolving societies without a leader, the average fraction of individuals who chose Green remained negligible (below 15%) even after the majority of citizens shifted their preferences toward Green (from period 15 onward). As a result, all Slow-evolving societies remained entrenched in the initially established norm, despite the presence of a 20% minority of individuals whose preferences changed at a faster pace. In contrast, 60% of the Fast-evolving societies (six out of ten) successfully transitioned to the new norm by the end of the game, despite a 20% minority of individuals exhibiting Slow-evolving preferences. The fraction of societies that adopted the new norm by the end of the game (the last five periods) is significantly different between the Slow-evolving and the Fast-evolving societies ($N = 10$ per treatment, one-sided Fisher's exact test, $P = 0.005$), treating each society as an independent observation (Slow-evolving: 0/10; Fast-evolving: 6/10).

Norm Change with a Leader. Comparing the societies where a leader was present with their respective baselines reveals that leaders successfully helped all societies abandon the old norm and coordinate on the new one (Fig. 3). Leaders had a very strong coordination power, as citizens followed their recommendations 96.5% of the time in Slow-evolving societies and 97.5% of the time in Fast-evolving societies (SI Appendix, Fig. S7). This finding

is remarkable, given the large group size, the public knowledge of the difference between the leader's and citizens' preference change patterns, and thus the potential for a conflict of interest in some periods. In Slow-evolving societies, the influence of the leader is highly significant, as all of them abandoned the Blue norm when a leader was present, while none was able to do so in the absence of a leader ($N = 10$ per treatment; one-sided Fisher's exact test, $P < 0.001$; with leader: 10/10; without leader: 0/10). In Fast-evolving societies, the difference between the leader treatment and the baseline is also significant ($N = 10$ per treatment, one-sided Fisher's exact test, $P = 0.043$; with leader: 10/10; without leader: 6/10). All Fast-evolving societies with a leader successively switched, whereas only 60% of them were able to achieve a norm change without a leader. The significance of the result remains when comparing the average proportion of individuals choosing Green between societies with a leader and their respective baselines ($N = 10$ per treatment, Mann-Whitney two-sided tests, $z = -3.329$ for SS-Baseline vs. FS-Baseline, $P < 0.001$; $z = -3.782$ for SS-Baseline vs. SS-FL, $P < 0.001$; $z = -2.874$ for FS-Baseline vs. FS-SL, $P = 0.004$). A regression analysis reported in SI Appendix, section C1 and Tables S3 and S4 confirms the results of the nonparametric tests.

Leadership Style and the Dynamics of Norms. We classified leaders as having a democratic style of leadership if they based their recommendations on majority preferences—recommending Green at most one period before the majority preference changed in Slow-evolving societies, or at most one period after the shift in Fast-evolving societies. This characterizes six out of twenty leaders overall (see societies 2 and 3 in the SS-FL treatment, and societies 3, 4, 5, and 6 in the FS-SL treatment in Fig. 3). The recommendations of these leaders for Green were

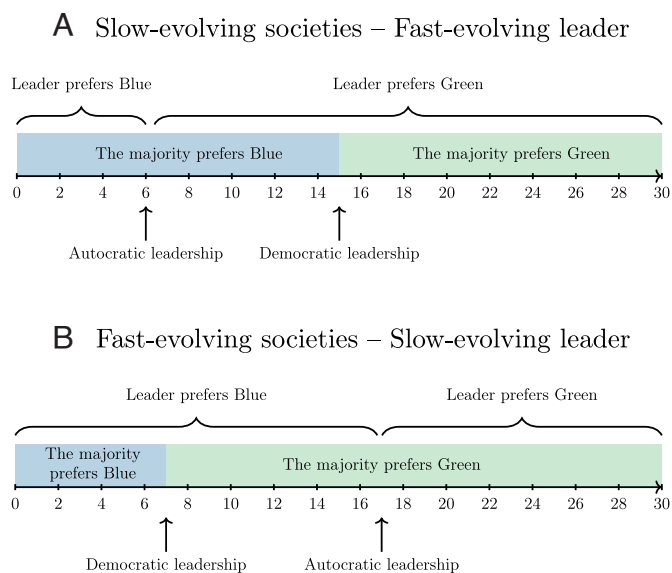


Fig. 1. Leadership styles. In the Slow-evolving Society–Fast-evolving Leader (SS-FL) treatment (A), the leader starts to prefer Green in period 6, whereas the majority only does so from period 15 onward. Conversely, in the Fast-evolving Society–Slow-evolving Leader (FS-SL) treatment (B), the leader starts to prefer Green only in period 17, whereas the majority does so from period 7. An autocratic-style leader should advocate for the adoption of the Green norm as soon as their own preference shifts—that is, in periods 6 in SS-FL and 17 in FS-SL. In contrast, a democratic-style leader should wait until the majority starts to prefer Green—that is, until periods 15 in SS-FL and 7 in FS-SL.

Slow-evolving societies - Baseline

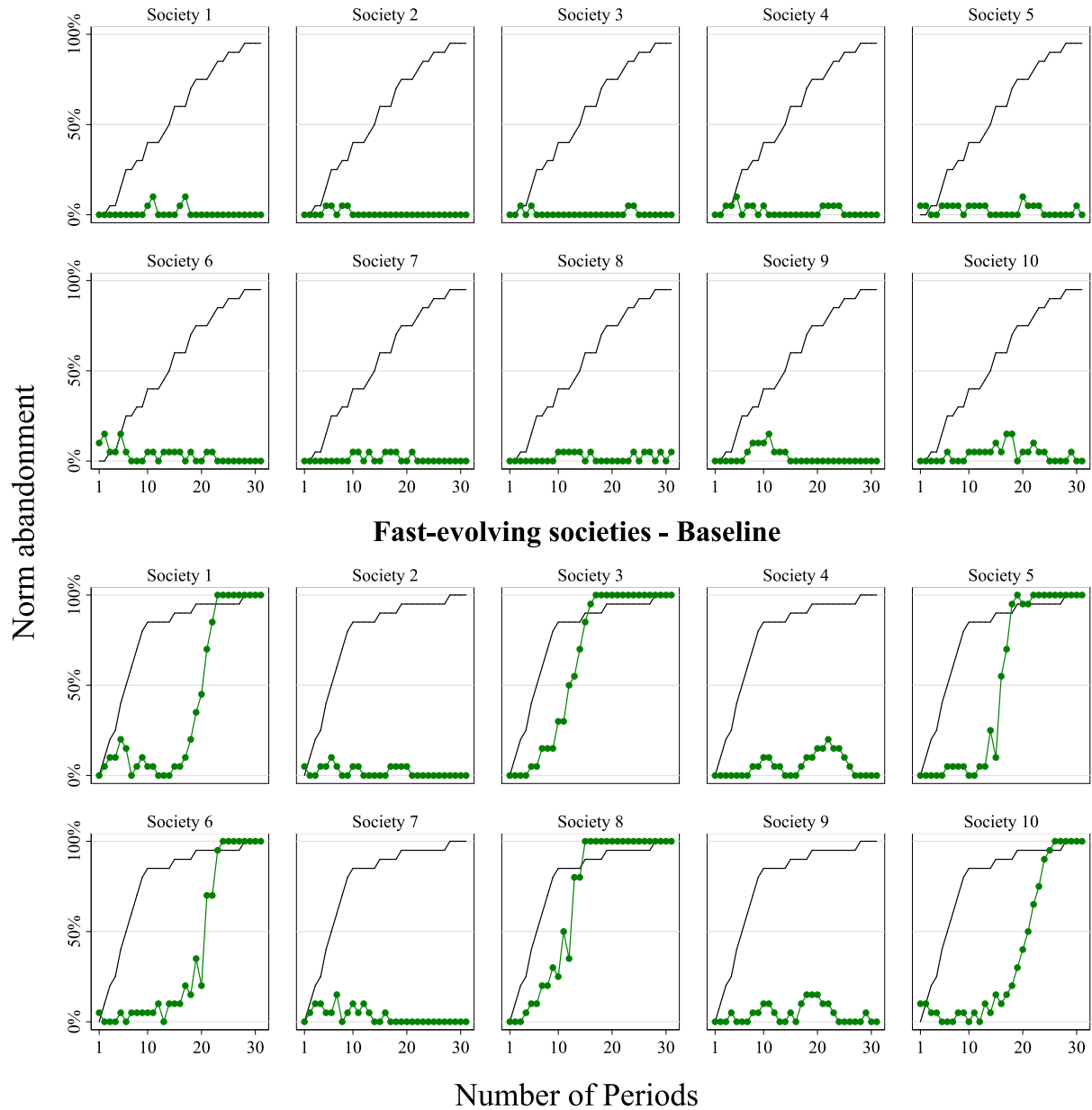


Fig. 2. Norm dynamics in each society, in the baseline treatments without leaders. The green line with circled markers represents the citizens' choices of Green and thus indicates norm abandonment. The solid black line shows the fraction of subjects who preferred to abandon the norm. A society is classified as having transitioned to a new norm if a majority of individuals coordinate on the Green norm during the final five periods of the game. In Slow-evolving societies 0 out of 10 abandoned the initially established norm. Only 60% of Fast-evolving societies (6 out of 10) abandoned the norm.

immediately adhered to by 72.5% and 78.75% of citizens in SS-FL and FS-SL, respectively. As a result, democratic leadership accelerated change when the society was Fast-evolving compared to when it was Slow-evolving. By contrast, we classified leaders as having an autocratic style of leadership if they based their recommendations on their personal preferences—advocating for Green at most one period after their personal preference shifted in Slow-evolving societies, or at most one period before their preference shift in Fast-evolving societies. The majority of leaders (11 out of 20) displayed an autocratic style (see societies 1, 4, 5, 6, 7, and 8 in the SS-FL treatment and societies 1, 2, 7, 8, and 9 in the FS-SL treatment in Fig. 3) (see *SI Appendix, section C2 and Table S5* on individual characteristics by leadership style).

The autocratic strategy was profitable for the leader only if at least two-thirds of the citizens adhered to the recommendation. Surprisingly, at least for the Slow-evolving societies, these leaders' first recommendation of choosing Green was immediately followed by a majority of, respectively, 60% (SS-FL) and 86% (FS-SL) of the citizens. Thus, autocratic leadership accelerated norm change in Slow-evolving societies with Fast-evolving leaders, and delayed it in Fast-evolving societies with Slow-evolving leaders. This finding is remarkable, considering that citizens could have anticipated the potential conflict of interest with the leader and thus could have chosen not to follow their recommendations.

Looking closer at the SS-FL treatment (*Top panel of Fig. 3*), we can distinguish i) situations in which Fast-evolving leaders adopting an autocratic style were immediately successful in

changing the norm, with more than two-thirds of the citizens following their recommendation (societies 4 and 5); and ii) situations in which they faced more resistance and had to revert to recommending the initial Blue norm for a couple of periods before making a second attempt to alter the prevailing norm (societies 1, 6, 7, and 8). Similarly, looking at the FS-SL treatment (*Bottom* panel of Fig. 3), we can see i) situations in which Slow-evolving leaders were successful in delaying the recommendation of Green, with more than two-thirds of the citizens waiting to implement this option (societies 1, 7, and 8); and ii) situations in which these leaders seemed to be exploring the consequences of their recommendations on their payoff before choosing the strategy that maximized their payoffs (societies 2 and 9).

We also identified a third type of leadership: the “compromising” leadership style in which leaders attempted to strike a compromise between the preferences of the majority and their own preferences. This was observed in three societies (9 and 10 in SS-FL and 10 in FS-SL).

Table 2 identifies the importance of the leader’s recommendation and leadership style, as well as citizens’ own preferences, in shaping citizens’ choice of Green, and it compares the influence of the two main leadership styles (autocratic vs. democratic).

When no leader is present, individual choices for Green largely reflect private preferences, as indicated by the positive and significant coefficient on Own preference for Green in both baseline conditions (SS-Baseline and FS-Baseline). The coefficient is smaller in Slow-evolving societies, consistent with the fact that these societies never succeed in transitioning to the Green norm.

The change in the size of the coefficient for Own preference for Green once a leader is introduced reflects how leadership reshapes the mapping from preferences to actions. In Slow-adjusting societies, preferences for Green become more predictive of behavior when a leader is present, as leader recommendations increasingly support the emerging preference for change over time. This reinforcement explains the modest increase in the coefficient relative to the baseline condition.

In Fast-adjusting societies, we observe the opposite pattern. Once leaders provide recommendations, participants partially substitute away from their private preferences toward this salient public signal, thereby reducing the weight placed on private preferences in determining behavior. Importantly, this does not imply that preferences become irrelevant or that their coefficients are directly comparable in magnitude to leadership effects, which are identified relative to the baseline recommendation (Democratic leaders recommending Blue). Rather, preferences remain positively associated with choosing Green while leadership adds a strong and independent coordinating influence that shifts decision weight from idiosyncratic preferences to a common focal point.

Regarding the influence of leadership style, Table 2 shows that when a leader recommended playing Green, its influence on citizens’ decisions was the same, irrespective of the leadership style (Wald test comparing the equality between the coefficients for a democratic vs. an autocratic leader recommending Green, $P = 0.807$ in SS-FL and $P = 0.550$ in FS-SL). The same is true when comparing recommendations to choose Blue in the SS-FL treatment. In the FS-SL treatment, however, an autocratic leader’s recommendation to play Blue had a negative and weakly significant effect on the choice of Green, compared to a democratic leader. These findings are important since they suggest that autocratic-style leaders may wield more power over

Table 2. Determinants of the choice of Green

	(1) SS-Baseline	(2) SS-FL	(3) FS-Baseline	(4) FS-SL
Own preference for Green	0.014** (0.005)	0.112*** (0.018)	0.403*** (0.107)	0.064*** (0.018)
Autocratic leader recommending Blue	–	0.017 (0.011)	–	–0.023* (0.012)
Democratic leader recommending Green	–	0.885*** (0.011)	–	0.898*** (0.021)
Autocratic leader recommending Green	–	0.887*** (0.019)	–	0.904*** (0.017)
Constant	0.011*** (0.004)	–0.014** (0.007)	0.001 (0.008)	0.022*** (0.007)
<i>N</i>	6,200	4,620	6,200	5,580
<i>R</i> ²	0.004	0.845	0.097	0.894

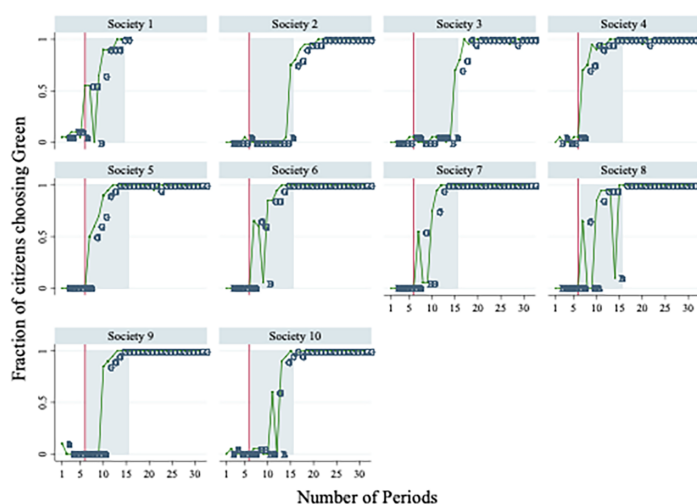
Notes: This table presents the estimates from random effects OLS models with SEs clustered at the society level (in parentheses). Columns (1) and (3) include the baseline conditions, whereas columns (2) and (4) correspond to the leader treatments. The dependent variable is a binary variable indicating whether a citizen chooses Green. The independent variables are binary variables representing the citizens’ preference for Green in the period and the leader’s recommendation in the period, depending on their leadership style. The reference category in the leader treatments is the recommendation of Blue by a leader adopting a democratic style. The difference in the number of observations across treatments comes from the fact that societies with a compromising leader are not included (one in FS-SL and two in SS-FL). *** $P < 0.01$ and * $P < 0.1$.

initiating or obstructing norm change than the majority of citizens.

Efficiency and Social Losses. Without a leader, societies forfeited approximately one-fourth of what they could have earned by choosing the color associated with the socially efficient outcome (*SI Appendix, Fig. S8*). Fast-evolving societies suffered as much as Slow-evolving societies from remaining stuck on the detrimental norm ($N = 10$ societies per treatment, two-sided Mann–Whitney U test, $z = -0.758$, $P = 0.448$), but significantly more from the penalties imposed when the two members of a pair failed to coordinate their actions ($N = 10$ societies per treatment, two-sided Mann–Whitney U test, $z = -3.331$, $P = 0.001$). Introducing a leader propelled both types of society to approach the socially efficient outcome. The presence of a leader significantly reduced the foregone payoffs due to adherence to the detrimental social norm in both types of society (Fast-evolving societies: $N = 10$ observations per treatment, two-sided Mann–Whitney U test, $z = -2.571$, $P = 0.010$; Slow-evolving societies: $N = 10$ observations for SS-FL and $N = 9$ for SS-FL where the observation for Society 1 is missing due to a technical z-tree crash, two-sided Mann–Whitney U test, $z = -3.707$, $P < 0.001$). It also significantly reduced miscoordination penalties in Fast-evolving societies ($N = 10$ observations per treatment, two sided Mann–Whitney U test, $z = 3.064$, $P = 0.002$), but not in Slow-evolving societies ($N = 10$ observations for SS-Baseline and $N = 9$ for SS-FL, two-sided Mann–Whitney U test, $z = 1.414$, $P = 0.157$).

Additionally, the average earnings of citizens and leaders depended on whether the leader adopted a democratic or autocratic leadership style (*SI Appendix, Table S6*). Leaders adopting an autocratic style earned significantly more on average than those using a democratic style: 835 vs. 719 ECU ($N = 10$ vs. 6 observations, two sided Mann–Whitney U test, $z = -3.254$, $P = 0.001$). In contrast, on average, citizens earned less under autocratic leadership than under democratic leadership: 781 vs. 817 ($N = 10$ vs. 6 observations, two sided Mann–Whitney U test, $z = 2.061$, $P = 0.039$).

Slow-evolving societies - Fast-evolving leaders



Fast-evolving societies - Slow-evolving leaders

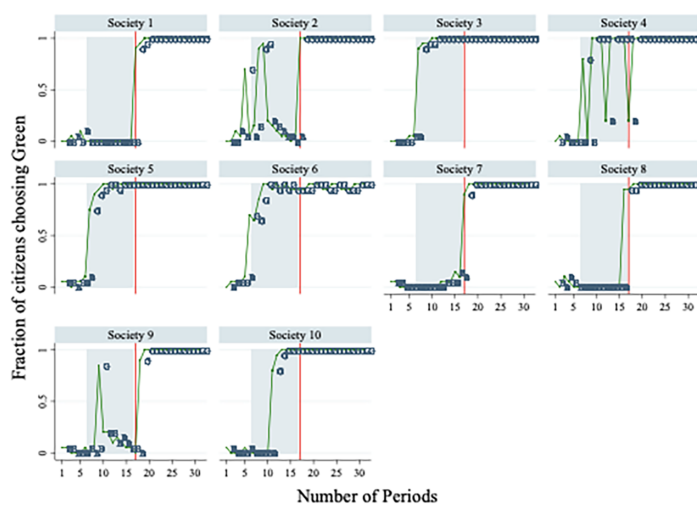


Fig. 3. Evolution of leaders' recommendations and of the proportion of citizens choosing Green in the leader treatments. The green solid line represents the fraction of citizens choosing the Green option over 31 periods in each society. The capital letters "B" (for Blue) and "G" (for Green) indicate the recommendation given by the leader to the group in each period. The vertical red line indicates the period when the leader's preference switched from Blue to Green. The blue-shaded area identifies the periods in which the leader's preference and those of the majority were misaligned. The figure shows that, in both treatments, all societies were able to transition to the more efficient norm. It also shows that six out of twenty leaders adopted a democratic leadership style: two in the SS-FL treatment (societies 2 and 3) and four in the FS-SL treatment (societies 3, 4, 5, and 6). The figure is incomplete for society 1 in the SS-FL treatment because the program crashed after period 14.

A Follow-Up Experiment with Peer-to-Peer Communication.

In our main experiment, leaders wielded strong power in coordinating citizens' expectations, allowing many to develop an autocratic strategy that prioritized their preferences over those of the citizens. However, in natural settings, several coordination mechanisms typically coexist. We conducted a follow-up experiment to test whether leaders relinquish their coordination power or alter their leadership style when citizens can communicate among themselves simultaneously to the leaders' recommendations.

An additional treatment ("FS-SL-Comm") was set up the same way as the FS-SL treatment, except that every five periods, starting from period 5, citizens were invited to simultaneously recommend a color (Blue/Green) to the other citizens. Communication was imperfect. Only 10 out of 20 messages sent by citizens were randomly selected and made public to all citizens, along with

the leader's message, before they could make a decision. Leaders knew the citizens could send messages, but they were unaware of their content.

Like in the FS-SL treatment, all societies in the new FS-SL-Comm treatment successfully transitioned to the Green norm before the end of the game (*SI Appendix, Fig. S9*). Interestingly, the norm change occurred, on average, in period 9.7 as opposed to period 13.4 in the absence of peer-to-peer communication ($N = 10$ observations per treatment, two-sided Mann-Whitney U test, $z = 1.724$, $P = 0.085$). This tendency toward accelerated transition could result from two forces: peer-to-peer communication itself and how leaders reacted to citizens communicating with each other.

Citizens behaved as expected: the proportion of recommendations for Green was 22% in period 5 (where the majoritarian preference was Blue), 73% in period 10, and more than 90%

from period 15. This large majority recommending Green from period 10, may have contributed to accelerating the transition to the Green norm.

Perhaps more determinant was the change in leadership style. Only one leader adopted autocratic leadership (society 5), as opposed to five such leaders in the FS-SL treatment. The frequency of democratic leadership was similar: three in FS-SL-Comm (societies 2, 6, and 9) vs. four in FS-SL. Most leaders (six out of ten: societies 1, 3, 4, 7, 8, and 10) tried to strike a compromise between their preferences and the citizens' preferences, whereas only one leader adopted such a compromising strategy in the FS-SL treatment (society 10). Overall, the difference in the frequency of each leadership style (autocratic, democratic, and compromising) between FS-SL-Comm and FS-SL is statistically significant ($N = 10$ observations per treatment, chi-squared test, $\chi^2(2) = 6.38, P = 0.041$). This suggests that, on average, leaders assigned a higher weight to the preferences of the majority when a concurrent coordination mechanism was available.

Discussion

The pressure on individuals to conform to others' behavior is a significant source of social norm inertia, even in settings where, at some point, the preferences of the majority of individuals would support abandoning an old norm that has become detrimental. We studied how heterogeneity in the speed of evolution of preferences in society affects the dynamics of social norms and how leaders manage normative transitions. In line with Andreoni et al. (18), we find that societies can remain stuck in conformity traps even when it is public knowledge that all individuals will eventually prefer Green. Using parameters calibrated to match their baseline environment—thereby reproducing the same tipping threshold (*SI Appendix, section B*)—we replicate their result, despite modeling preference change through a linear rather than constant probability process. In our experiment, all Slow-evolving societies exhibited such a conformity trap, despite the presence of a minority of individuals whose preferences changed earlier but who ultimately failed to act as effective instigators of change. This suggests that the mere presence of a few forward-looking individuals who anticipate society's future preferences is insufficient to instigate change. Conversely, in Fast-evolving societies, where the majority of citizens' preferences evolved at a faster pace, the likelihood of a norm change was higher, despite the presence of a minority with sticky preferences. This suggests that, in a Fast-evolving society, it is possible to escape the conformity trap, and a minority of individuals with slow-evolving preferences cannot block this impetus for change, provided that the majority generates enough instigators of change spontaneously to initiate the process (see *SI Appendix, section C3*, for an analysis of instigators of change). However, even in such a fast-changing setting, only 60% of the societies successfully transitioned to the new social norm by the end of the game, and this was obtained at the cost of higher frictions compared to Slow-evolving societies. This suggests that many fast-evolving individuals, even when aware that they likely represented the majority, believed that they had little hope of expediting change through their behavior. Despite the attempts of instigators to change the norm, a powerful coordination mechanism was needed to align social expectations and persuade individuals to behave according to a new norm.

Due to their prominence, leaders making public recommendations served as a powerful coordination mechanism in

shaping social expectations, irrespective of the type of society or their own type. Leaders increased the probability of successful norm change by 100 percentage points in the Slow-evolving societies (from none of the 10 societies in the baseline to all of the 10 societies with a leader) and by 40 percentage points in the Fast-evolving societies (from six of the 10 societies in the baseline to all of the 10 societies with a leader). All Fast-evolving leaders successfully steered Slow-evolving societies out of the conformity trap, sometimes even before the majority of citizens changed their preferences. Even Slow-evolving leaders, while often attempting to delay the norm change, helped all societies abandon the old norm, sometimes even before the leader's preference switched. The effect of leadership is at least as large as, and in some cases substantially larger than the most powerful coordination interventions in Andreoni et al. (18): While in their baseline (TT-43), none of the societies switched to the new norm, the Preference Poll treatment—one of their most effective interventions—led 5 out of 6 societies (83%) to successfully switch, while the Incentive for Instigators treatment resulted in norm change in 3 out of 6 societies (50%).

Heterogeneity in the leadership style influenced the process of norm change. We observed that a democratic leadership style was adopted less frequently than an autocratic style, as it imposes relatively higher costs on the leader and could be considered a form of "leadership-by-sacrifice" (31). Autocratic leadership was as successful as democratic leadership in swaying citizens' behavior, and it unexpectedly granted leaders a power exceeding that of the majority. However, it was less welfare-enhancing compared to democratic leadership.

Leaders who adopted a democratic leadership style in our experiment exhibited higher risk aversion than those who adopted an autocratic style. Importantly, adopting a democratic leadership style should not be interpreted as a fixed personality trait but rather as reflecting a strategic choice to accommodate the majority. This choice may be driven by coordination concerns, risk considerations, and a willingness to respect majority preferences. Nonetheless, the influence of risk attitudes should be interpreted with caution given the limited number of leaders in our study.

Autocratic leadership, where leaders seek to align social expectations with their own preferences, is, however, more difficult to sustain when other coordination mechanisms are available within the society. Our follow-up experiment showed that peer-to-peer communication of preferences regarding collective action, in addition to the leader's recommendations, rebalanced the power relationship within the society. It increased leaders' willingness to seek a compromise between their preferences and those of the majority of citizens. An important policy implication of these findings is that if coordination of social expectations is managed from the top, facilitating successful transitions to new social norms, peer-to-peer communication may help redistribute the costs and benefits of normative transitions.

More research is needed to understand the dynamics of social norms and how they interact with the change in preferences. Indeed, while we exogenously imposed the pattern of preference changes, many examples worldwide highlight the crucial challenge of instigating changes in preferences before effectively coordinating social expectations. We have identified cases in which the leader was able to move society to a tipping point before the majority shifted preference, but in a setting where everyone could anticipate that the majority preference would change at some point.

Further research is also needed to better understand the limits of leaders' ability to coordinate normative expectations and

behavior. In particular, it would be valuable to identify the maximum share of individuals who never change preferences beyond which a leader's coordination power breaks down. Exploring how different leadership forms shape normative dynamics in societies would be a natural extension of our study, for example, through "leading-by-example" or by allowing citizens to vote for selecting and renewing their leader. Studying leadership under alternative incentive structures also is important. Specifically, when leaders benefit disproportionately from a new norm relative to followers, how does this affect their communication strategies and their capacity to trigger normative change?

Finally, it would be important to examine how competition between leaders, potentially holding different preferences, would affect leadership style and its impact on the pattern of norm change in the society.

Materials and Methods

We recruited 1,030 participants among students, mainly from local business and engineering schools, using the H-Root online recruitment platform (46). The mean age was 23 y, and 50% of participants were females. We ran a total of 50 sessions with 10 sessions per treatment (eight at Gate-Lab in Lyon and two at Laboratoire d'Economie Appliquée de Grenoble, France, because of subject-pool size constraints (see *SI Appendix, Tables S7 and S8* for statistics on the sociodemographic characteristics of the participants in the role of citizens and leaders, respectively. *SI Appendix, Tables S9 and S10* provide similar information for the follow-up experiment). The Review Board of the Groupe d'Analyse et de Théorie Economique at the University of Lyon reviewed and approved the procedures (2021-12). Participants gave their consent and were informed that they could opt out at any time. We programmed the experiment with the z-Tree software (47).

Upon arrival, participants received written instructions (also read aloud) about the game (*SI Appendix*), and answered a comprehension questionnaire. Once all participants answered correctly, we started the game on computer terminals. Subjects participated in only one of the five treatments, either in the role of a citizen or in the role of a leader, with roles kept fixed. Throughout the experiment, we used neutral language. For example, we never used the term "leader" (instead, we talked of "type C" participant), and to identify a change in preferences, we stated that participants moved from being identified as "type A" to "type B." Then, participants played a rule-following game (48) and answered a sociodemographic questionnaire. In this questionnaire, they self-reported their risk attitude, patience, and general trust on a scale from 1 (lowest) to 10 (highest), as well as their political orientation on a scale from 1 (strongly liberal orientation) to 10 (strongly conservative orientation). Finally, we collected information on gender, age, level of education, and school, and included a measure of nonconformism.

At the end of the experiment, we paid participants in cash and in private in a separate room. They earned on average €29.7 (s.d. = €4.2) in the main experiment and €32 (s.d. = €3.51) in the follow-up treatment, including a €5 show-up fee, for a two-hour session on average.

Data, Materials, and Software Availability. Data and programs have been deposited in Open Science Framework (<https://doi.org/10.17605/OSF.IO/JPW7X>) (49).

ACKNOWLEDGMENTS. We are grateful to G. d'Adda, E. Fatas, S. Gächter, M. Micalel, N. Nikiforakis, S. Siegenthaler, the Editor and two anonymous reviewers, and participants at the Association Française d'Economie Expérimentale conference in Lyon, the Economic Science Association (ESA) World meeting in Lyon, the North-American ESA meeting in Santa Barbara, the Africa ESA meeting in Cape Town, the Asia-Pacific ESA meeting in Singapore, the European ESA meeting in Helsinki, the 7th Annual Colfer Conference on Behavioral Economics in Tel-Aviv, the Society for the Advancement of Behavioral Economics - International Association for Research in Economic Psychology conference in Nice, the Maastricht Behavioral and Experimental Economics Symposium in Maastricht, the 23rd Louis-André Gérard-Varet Conference in Marseille, the Tiber Symposium on Psychology and Economics in Tilburg, the Asia Meeting of the Econometric Society in Singapore, the workshop on Behavioral and Experimental Economics for Innovative Policy-Making in Montreal, the "Unraveling the Dynamics of Social Change" workshop at New York University Abu Dhabi, and the Shandong University Golden Autumn Cooperation Workshop in Jinan for helpful comments. We also benefited from comments when presenting the paper in seminars at the Center for Research in Experimental Economics and Political Decision Making (Amsterdam), briq (Bonn), the Institute of Social and Economic Research (Osaka), Lille Economie Management (Lille), Middlebury College (Middlebury Virtual Experimental Economics Seminar), New-York University Abu Dhabi, and Libera Università Internazionale degli Studi Sociali Guido Carli (Roma). We are thankful to Q. Thevenet for providing programming assistance and S. Robin for allowing us to conduct experimental sessions at the Laboratoire d'Economie Appliquée de Grenoble. Portions of the paper were developed from the thesis of J.K. This research has benefited from the financial support of IDEXLYON from Université de Lyon (project Institutional Design and Economic Preferences: Theory and experiments) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005), the project DECISION of the French National Research Agency (ANR-19-CE26-0019), and the LABEX CORTEX (ANR-11-LABX-0042) of Université Claude Bernard Lyon 1, within the Programme "Investissements d'Avenir" (2019-ANR-LABX-02) operated by the French National Research Agency (ANR). J.K. acknowledges support under ERC-AdG 101020453 PRINCIPLES.

Author affiliations: ^aCNRS, Université Lumière Lyon 2, Université Jean-Monnet Saint-Etienne, Emlyon Business School, Groupe d'Analyse et de Théorie Economique, Lyon 69007, France; and ^bCentre for Decision Research and Experimental Economics, University of Nottingham, Nottingham NG7 2QL, United Kingdom

1. C. Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (Oxford University Press, 2017).
2. C. Bicchieri, C. Rovelli, Evolution and revolution: The dynamics of corruption. *Ration. Soc.* **7**, 201-224 (1995).
3. G. Sparkman, L. Howe, G. Walton, How social norms are often a barrier to addressing climate change but can be part of the solution. *Behav. Public Policy* **5**, 528-555 (2021).
4. E. Fehr, U. Fischbacher, S. Gächter, Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1-25 (2002).
5. L. Balafoutas, N. Nikiforakis, B. Rockenbach, Direct and indirect punishment among strangers in the field. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15924-15927 (2014).
6. L. Balafoutas, N. Nikiforakis, B. Rockenbach, Altruistic punishment does not increase with the severity of norm violations in the field. *Nat. Commun.* **7**, 1-6 (2016).
7. Z. H. Garfield *et al.*, Norm violations and punishments across human societies. *Evol. Hum. Sci.* **5**, e11 (2023).
8. C. Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge University Press, 2005).
9. H. P. Young, The evolution of social norms. *Annu. Rev. Econ.* **7**, 359-387 (2015).
10. L. Novak, Persistent norms and tipping points: The case of female genital cutting. *J. Econ. Behav. Organ.* **177**, 433-474 (2020).
11. S. Gulesci *et al.*, A Stepping Stone Approach to Norm Transitions. *Am. Econ. Rev.* **115**, 2237-2266 (2025).
12. M. Granovetter, Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420-1443 (1978).
13. M. Granovetter, R. Soong, Threshold models of diffusion and collective behavior. *J. Math. Sociol.* **9**, 165-179 (1983).
14. C. Bicchieri, A. Funcke, Norm change: Trendsetters and social structure. *Soc. Res.* **85**, 1-21 (2018).
15. D. Centola, J. Becker, D. Brackbill, A. Baronchelli, Experimental evidence for tipping points in social convention. *Science* **360**, 1116-1119 (2018).
16. C. Efferson, S. Vogt, E. Fehr, The promise and the peril of using social influence to reverse harmful traditions. *Nat. Hum. Behav.* **4**, 55-68 (2020).
17. G. Andrighetto, E. Vriens, A research agenda for the study of social norm change. *Philos. Trans. R. Soc. A* **380**, 20200411 (2022).
18. J. Andreoni, N. Nikiforakis, S. Siegenthaler, Predicting social tipping and norm change in controlled experiments. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014893118 (2021).
19. E. L. Ferrara, A. Chong, S. Duryea, Soap operas and fertility: Evidence from Brazil. *Am. Econ. J. Appl. Econ.* **4**, 1-31 (2012).
20. T. Lane, D. Nosenzo, S. Sonderegger, Law and norms: Empirical evidence. *Am. Econ. Rev.* **113**, 1255-1293 (2023).
21. F. Casoria, F. Galeotti, M. C. Villeval, Perceived social norm and behavior quickly adjusted to legal changes during the covid-19 pandemic in France. *J. Econ. Behav. Organ.* **190**, 54-65 (2021).

22. R. Galbiati, E. Henry, N. Jacquemet, M. Lobeck, How laws affect the perception of norms: Empirical evidence from the lockdown. *PLoS One* **16**, e0256624 (2021).
23. M. E. Tankard, E. L. Paluck, The effect of a supreme court decision regarding gay marriage on social norms and personal attitudes. *Psychol. Sci.* **28**, 1334–1344 (2017).
24. L. Bursztyn, G. Egorov, R. Enikolopov, M. Petrova, Social media and xenophobia: Evidence from Russia (CEPR Discussion Paper, 14877, CEPR Press, Paris & London, 2020).
25. D. Acemoglu, M. O. Jackson, History, expectations, and leadership in the evolution of social norms. *Rev. Econ. Stud.* **82**, 423–456 (2015).
26. L. Bursztyn, G. Egorov, S. Fiorin, From extreme to mainstream: The erosion of social norms. *Am. Econ. Rev.* **110**, 3522–3548 (2020).
27. M. Abel, W. Brown, Prosocial behavior in the time of covid-19: The effect of private and public role models. *J. Behav. Exp. Econ.* **101**, 101942 (2022).
28. K. Müller, C. Schwarz, From hashtag to hate crime: Twitter and antiminority sentiment. *Am. Econ. J. Appl. Econ.* **15**, 270–312 (2023).
29. N. Aizenman, The power of example: Corruption spurs corruption. *Am. Econ. J. Appl. Econ.* **13**, 230–257 (2021).
30. N. Aizenman, T. Cavalcanti, D. Da Mata, More than words: Leaders' speech and risky behavior during a pandemic. *Am. Econ. J. Econ. Policy* **15**, 351–371 (2023).
31. B. E. Hermalin, Toward an economic theory of leadership: Leading by example. *Am. Econ. Rev.* **88**, 1188–1206 (1998).
32. J. Potters, M. Sefton, L. Vesterlund, Leading-by-example and signaling in voluntary contribution games: An experimental study. *Econ. Theory* **33**, 169–182 (2007).
33. M. Komai, M. Stegeman, B. E. Hermalin, Leadership and information. *Am. Econ. Rev.* **97**, 944–947 (2007).
34. G. Eisenkopf, Words and deeds-experimental evidence on leading-by-example. *Leadersh. Q.* **31**, 101383 (2020).
35. D. J. Cooper, J. R. Hamman, R. A. Weber, Fool me once: An experiment on credibility and leadership. *Econ. J.* **130**, 2105–2133 (2020).
36. S. Gächter, E. Renner, Leaders as role models and 'belief managers' in social dilemmas. *J. Econ. Behav. Organ.* **154**, 321–334 (2018).
37. L. Hoefl, W. Mill, A. Vostroknutov, Normative acceptance of power abuse. *Kyklos* **78**, 1375–1386 (2025).
38. G. d'Adda, D. Darai, N. Pavanini, R. A. Weber, Do leaders affect ethical conduct? *J. Eur. Econ. Assoc.* **15**, 1177–1213 (2017).
39. J. Antonakis, G. d'Adda, R. A. Weber, C. Zehnder, "Just words? just speeches?" on the economic value of charismatic leadership. *Manag. Sci.* **68**, 6355–6381 (2022).
40. R. K. Wilson, C. M. Rhodes, Leadership and credibility in n-person coordination games. *J. Confl. Resolut.* **41**, 767–791 (1997).
41. R. Weber, C. Camerer, Y. Rottenstreich, M. Knez, The illusion of leadership: Misattribution of cause in coordination games. *Organ. Sci.* **12**, 582–598 (2001).
42. J. Brandts, D. J. Cooper, R. A. Weber, Legitimacy, communication, and leadership in the turnaround game. *Manag. Sci.* **61**, 2627–2645 (2015).
43. S. G. Sahin, C. Eckel, M. Komai, An experimental study of leadership institutions in collective action games. *J. Econ. Sci. Assoc.* **1**, 100–113 (2015).
44. E. Reuben, K. Timko, On the effectiveness of elected male and female leaders and team coordination. *J. Econ. Sci. Assoc.* **4**, 123–135 (2018).
45. P. J. Grossman, C. Eckel, M. Komai, W. Zhan, It pays to be a man: Rewards for leaders in a coordination game. *J. Econ. Behav. Organ.* **161**, 197–215 (2019).
46. O. Bock, I. Baetge, A. Nicklisch, Hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev.* **71**, 117–120 (2014).
47. U. Fischbacher, z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **10**, 171–178 (2007).
48. E. O. Kimbrough, A. Vostroknutov, A portable method of eliciting respect for social norms. *Econ. Lett.* **168**, 147–150 (2018).
49. F. Galeotti, J. Krutaj, M. C. Villeval, The power of leadership in changing social norms in heterogeneous societies. OSF. <https://doi.org/10.17605/OSF.IO/JPW7X>. Deposited 16 February 2026.

PNAS



Supporting Information for

The power of leadership in changing social norms in heterogeneous societies

Fabio Galeotti, Jona Krutaj, Marie Claire Villeval

Corresponding author: Marie Claire Villeval.

E-mail: villeval@gate.cnrs.fr

This PDF file includes:

- Figs. S1 to S9
- Tables S1 to S16
- SI References
- Instructions

A. Materials and methods.

A.1. Payoff functions.

Citizens' payoff function. In the experimental instructions, we used a neutral language to identify preferences: participants were identified as "type A" (i.e., preferring Blue) or "type B" (i.e., preferring Green). We induced preferences over the Blue and Green options by assigning a higher payoff, $v_H = 30$ Experimental Currency Units (ECU), if citizens chose one color, and a lower payoff, $v_L = 20$ ECU, if they chose the other color.

Let a_i and a_j denote the actions (i.e., the choices of a color) of citizens i and j respectively, and let n_i and n_j denote the norms preferred by citizens i and j respectively, where j is the partner of i in a given period. The payoff function for citizen i in that period is defined as follows:

$$\pi_i = \begin{cases} v_H & \text{if } a_i = a_j \quad \& \quad a_i = n_i \\ v_L & \text{if } a_i = a_j \quad \& \quad a_i \neq n_i \\ v_H - p \times Nb_{others} & \text{if } a_i \neq a_j \quad \& \quad a_i = n_i \\ v_L - p \times Nb_{others} & \text{if } a_i \neq a_j \quad \& \quad a_i \neq n_i \end{cases}$$

where Nb_{others} represents the number of citizens choosing the opposite action compared to i . Failure of i and j to coordinate triggers a penalty equal to a fixed rate ($p = 4$) multiplied by the number of citizens choosing the other action.

Table S1 displays how citizens' payoffs were calculated in the experiment. The upper part of Table S1 shows the payoffs resulting from the choices of two citizens who were matched together at the beginning of the game. All citizens were induced to prefer Blue at the beginning of the game. They received a higher payoff if they coordinated with their partner on playing Blue rather than Green: (BB= v_H, v_H) vs. (GG= v_L, v_L). The cost of miscoordination was $p \times (g_t)$ for an individual choosing Blue and $p \times (1 - g_t)$ for an individual choosing Green, where p was the fixed penalty parameter and g_t the proportion of citizens choosing Green.

The middle part of the table represents the intermediate situation in which one citizen in the pair preferred Blue whereas the other one preferred Green. The bottom part of the table displays the payoffs at the end of the game when both preferred Green.

Table S1. Citizens' payoffs

Both citizens prefer Blue	Blue	Green
Blue	v_H, v_H	$v_H - p \times g_t, v_L - p \times (1 - g_t)$
Green	$v_L - p \times (1 - g_t), v_H - p \times g_t$	v_L, v_L
1 prefers B/ 1 prefers G	Blue	Green
Blue	v_H, v_L	$v_H - p \times g_t, v_H - p \times (1 - g_t)$
Green	$v_L - p \times (1 - g_t), v_L - p \times g_t$	v_L, v_H
Both citizens prefer Green	Blue	Green
Blue	v_L, v_L	$v_L - p \times g_t, v_H - p \times (1 - g_t)$
Green	$v_H - p \times (1 - g_t), v_L - p \times g_t$	v_H, v_H

Notes: These matrices represent the citizens' payoffs in three situations, depending on the preferences of interacting pair members and their respective choices, Blue or Green. p is the fixed penalty for miscoordination with the partner. $1 - g_t$ is the proportion of citizens who chose Blue. g_t is the proportion of citizens who chose Green. In the experiment, $v_H = 30$ Experimental Currency Units (ECU), $v_L = 20$ ECU, and $p = 4$.

At the end of each period t , citizens received feedback on the choice of their paired member in period t , and their own earnings in $t - 1$. Citizens never observed the preferences of the other members of society, nor were they ever informed about how many people had changed preferences in a given period. They were informed about the choice of the others after a delay of one period. This reflects the fact that people in the real world do not have immediate access to information about the choices made by everyone in society. However, since the probability of changing preferences was publicly known, subjects should have anticipated that everyone would eventually prefer the Green option. At some point, societies benefited from abandoning the Blue norm in favor of the Green norm. This means that from a certain period onward, citizens received a higher payoff if they coordinated with the others on Green as opposed to Blue: (BB= v_L, v_L) vs. (GG= v_H, v_H).

Leaders' payoff function. The leaders' payoff function was defined as follows:

$$\pi_{Leader} = \begin{cases} 900 - 70 \times Nb_{Incongruent} & Nb_{Congruent} \geq Nb_{Incongruent} \\ 400 - 20 \times Nb_{Congruent} & Nb_{Incongruent} \geq Nb_{Congruent} \end{cases}$$

where $Nb_{Incongruent}$ ($Nb_{Congruent}$) identifies the number of citizens who choose a color that is incongruent (congruent) with the leader's preference.

This payoff function, illustrated in Fig. S2 and known to all participants, indicates that the leader was incentivized to induce the society to conform to her preferred norm. Indeed, the leader received a significantly higher payoff if *all* the citizens chose her preferred norm than if they *all* chose the alternative norm (900 vs. 400 ECU). When the majority of the citizens chose a color congruent with the leader's preference, the leader started with a base payoff of 900 ECU and suffered a fixed penalty of 70 ECU for each citizen who chose an action incongruent with the leader's preference. The fixed penalty for each individual who deviated from the choice of the majority depended on whether these trendsetters supported the same norm as the leader or not. The fixed penalty for the leader was 20 (70) ECU when the trendsetters chose the norm the leader preferred (did not prefer). As a consequence, the leader's payoff decreased and reached a minimum (200 ECU) when the society was completely divided, with half of the citizens choosing Blue and the other half choosing Green.

With this payoff function, we intended to capture a setting where the leader benefits if the normative conflict in society is minimized. To avoid the minimum payoff, the leader should stop recommending her favorite option if she believes fewer than two-thirds of the citizens will follow her recommendation. To illustrate this, suppose that the leader prefers Blue. If she recommends Blue and 13 citizens out of 20 follow, she earns $900 - 70 \times (20 - 13) = 410$ ECU. However, if one less person chooses Blue, then the leader's payoff would be reduced to 340 ECU, which is less than what she would get if she recommended the alternate option and everyone followed her (400 ECU).

In the experiment, we paid citizens the sum of their payoffs over the 31 periods (as in (1)), and we randomly selected one period to determine the earnings of the leaders, with 100 ECU = 3 Euros. We aimed to make the monetary consequences of each leader's decision substantial.

A.2. Additional information on preference changes. To keep the dynamics of preference changes constant across sessions within the same type of society, we predetermined when each participant would change preference in each society and applied the same distribution of preference changes in each session of a given condition. To avoid deception, we informed participants that the exact period during which they would change type had already been determined before the beginning of the session. We also told them that the selection was done independently for each member of the group, based on simulations that reflected a probability of change that increased with each period by 6 [0.6] percentage points for 16 participants and by 0.6 [6] percentage points for 4 participants. We did so by generating, for each type of society, 1000 draws that followed the theoretical distribution of probabilities for this society (see Fig. S3). Then, we chose the distribution that minimized the sum of squared errors between the

expected distribution and the actual distribution. Alternatively, we could have used a Kolmogorov-Smirnov test to select the distribution that best represents the theoretical distribution of probabilities. Both methods led to the selection of only one distribution for Slow-evolving societies. However, the Kolmogorov-Smirnov criterion selected nine different distributions for Fast-evolving societies, while the minimization of the sum of squared errors selected only one. According to the selected distribution (see Fig. S3), the majority of citizens preferred Green from period 7 in Fast-evolving societies and from period 15 in Slow-evolving societies. Thus, for the majority, it became beneficial to coordinate on the Green norm from period 7 until the end of the game in Fast-evolving societies and from period 15 in Slow-evolving societies. Moreover, at least 80% of the citizens preferred Green in both the Fast-evolving and the Slow-evolving societies in the last ten periods and at least 90% in the last five periods.

Following the same procedure, the switching period of the leader's preferences was selected by taking the average of 1000 draws following the respective probabilities. The average switching period was 5.8 for a Fast-evolving leader and 16.6 for a Slow-evolving leader.

B. Tipping threshold model. In this section, we summarize the tipping threshold model proposed by Andreoni, Nikiforakis, and Siegenthaler (2021) (1), which can also be used to analyze norm change in the presence of heterogeneity in the dynamics of preference changes. We then discuss how the introduction of leadership can shift the system relative to the predictions of Andreoni, Nikiforakis, and Siegenthaler (2021).

Their theoretical model adopts a rational-choice framework and assumes that each individual is characterized by a threshold f_i . This threshold represents the proportion of other individuals that individual i requires to have deviated from the existing norm before i would also do so. Specifically, this threshold is the point at which the expected payoff from switching to the new norm exceeds the anticipated payoff from sticking with the old norm.

This individual threshold, f_i , is derived by comparing the expected benefits from choosing Green and Blue. The expected benefit from sticking with Blue depends on the net benefit that the option Blue provides (v_{Blue}) and the expected cost of miscoordinating on Green, which is equal to the probability of miscoordinating (equivalent to the proportion of other individuals choosing Green, symbolized by f_{Green}) multiplied by the penalty incurred ($P \times f_{Green}$). P represents the maximum penalty incurred when all the other individuals choose the opposite color. The formula for calculating this maximum penalty is $P = p \times (N - 1)$, where N represents the total number of citizens in society and p the penalty for each citizen choosing the opposite color. The expected payoff from choosing Blue is written as follows:

$$E[\pi_i(Blue)] = v_{Blue} - P f_{Green}^2 \quad [1]$$

Symmetrically, the expected payoff from choosing Green depends on the benefit of choosing Green (v_{Green}) and the expected cost of miscoordinating ($f_{Blue}(P \times f_{Blue})$), where f_{Blue} is the proportion of individuals who choose Blue. It also depends on the expected benefit of expediting a change in the norm, which depends on the net benefit of changing the norm, $v = v_{Green} - v_{Blue}$, multiplied by self-efficacy to change the norm (γ_i). The parameter γ_i captures the belief of individual i that their actions will accelerate the change in norm. The higher the self-efficacy parameter, the sooner the individual believes everyone else will follow, and the higher the expected benefit from choosing Green. Such a component is not present in $E[\pi_i(Blue)]$ because choosing Blue does not change the norm. The expected payoff from choosing Green is written as follows:

$$E[\pi_i(Green)] = v_{Green} - P f_{Blue}^2 + \gamma_i v. \quad [2]$$

Equating Eq. (1) and Eq. (2), and solving for f_{Green} gives the individual threshold f_i (note that $f_{Blue} = 1 - f_{Green}$):

$$f_i = \frac{1}{2} - \frac{1}{2} \left[\frac{v}{P} (1 + \gamma_i) \right] \quad [3]$$

f_i indicates the point at which an individual becomes indifferent between choosing Blue and Green. The individual threshold depends on: (i) the net benefit from choosing one's preferred color, $v = v_{Green} - v_{Blue}$, (ii) the maximum penalty from failing to conform to the norm, P , and (iii) the expectation about how one's deviation from the norm affects the likelihood of a successful change in the norm, which is captured by γ_i . It is assumed that $\gamma_i \sim N(\mu, \sigma)$. The formula for the individual threshold indicates that individuals are more inclined to switch to the new norm earlier when the net benefit (v) is higher, the maximum penalty (P) is lower, and they have a stronger belief in self-efficacy (γ_i).

The tipping threshold of society is the point where the number of people deviating from the norm is large enough such that even individuals who do not believe they can expedite change ($\gamma_i = 0$) start choosing Green. Thus, the societal tipping threshold is given by the following formula:

$$f_T = \frac{1}{2} - \frac{1}{2} \left(\frac{v}{P} \right) \quad [4]$$

Incorporating the societal tipping threshold into the individual threshold by substituting Eq. (4) into Eq. (3) yields the following formula for the individual threshold:

$$f_i = f_T - \frac{1}{2} \frac{v \gamma_i}{P} \quad [5]$$

The main assumption of this tipping threshold model is that, given a distribution of f_i , if the proportion of individuals who have abandoned the norm at the beginning of period t is $q(t)$, then in period $t + 1$, all individuals with a threshold $f_i \leq q(t)$ will abandon the norm as well.

The probability that an individual will choose Green in a given period t is characterized as follows:

$$\mathbb{P}(\text{choice}_t = \text{Green}) = \mathbb{P}(f_i \leq f_{Green,t-1}) = \mathbb{P}\left(f_T - \frac{1}{2} \frac{v \gamma_i}{P} \leq f_{Green,t-1}\right) \quad [6]$$

Rearranging the terms yields:

$$\mathbb{P}(\text{choice}_t = \text{Green}) = \mathbb{P}\left(\frac{f_T - f_{Green,t-1}}{0.5v/P} \leq \gamma_i\right) \quad [7]$$

Recall that $\gamma_i \sim N(\mu, \sigma)$. After calling $\tilde{\gamma} = \frac{f_T - f_{Green,t-1}}{0.5v/P}$ and standardizing γ_i , we obtain:

$$\mathbb{P}(\text{choice}_t = \text{Green}) = \mathbb{P}\left(\frac{\tilde{\gamma} - \mu}{\sigma} \leq z\right) = \mathbb{P}\left(z \leq \frac{\mu - \tilde{\gamma}}{\sigma}\right) = \Theta\left(\frac{\mu - \tilde{\gamma}}{\sigma}\right) \quad [8]$$

where Θ is the cumulative distribution function of the standard normal distribution. This is equivalent to a Probit model where $\tilde{\gamma}$ is the independent variable. The coefficient of this variable provides an estimate of $-1/\sigma$, while the intercept of the model provides an estimate of $\frac{\mu}{\sigma}$. From these estimates, it is possible to derive the distribution of γ_i .

In our setting, we use the same v and P parameters as in the baseline of Andreoni, Nikiforakis, and Siegenthaler (2021); consequently, the tipping threshold is also the same and equal to $f_T = 0.43$ (as derived from Eq. (4)). The key difference relative to Andreoni, Nikiforakis, and Siegenthaler (2021) lies in the way preferences evolve over time: in our design, preference change follows a linear probability process, whereas in their setting it follows a constant probability process. As a result, any difference between our societies and the baseline of Andreoni, Nikiforakis, and Siegenthaler (2021) arises from the differences in the likelihood of reaching the tipping threshold, which is shaped by the distribution of γ_i . Estimating the distribution of γ_i using the data from our Slow-evolving societies gives $\gamma_i \sim N(1.73, 2.08)$. Based on the distribution of γ_i , we can compute the distribution of the individual tipping thresholds for Slow-evolving societies (solid line in Fig. S4). Almost all members of our Slow-evolving societies had a positive switching threshold and were thus reluctant to instigate change. This is in line with the observation that none of the Slow-evolving societies escaped the conformity trap. In contrast, when using the data from our Fast-evolving societies, $\gamma_i \sim N(2.08, 2.32)$. Computing the distribution of the individual tipping thresholds for Fast-evolving societies (dashed line in Fig. S4) shows that this distribution is slightly shifted to the left. A few more individuals had a threshold equal to or lower than zero, indicating a willingness to instigate change

independently of the number of other people who chose the Green norm. This implies that a change in the norm was more likely to occur. We found that six out of ten Fast-evolving societies were able to abandon the Blue norm without external intervention.

We now illustrate how leadership can affect the aforementioned theoretical framework. We argue that leadership affects behavior by shifting individuals' expectations about others' behavior. In the absence of leadership, these expectations simply reflect the proportion of individuals who had abandoned the norm at the beginning of the previous period (i.e., the only information available to citizens at the start of a new period). In the presence of a leader, their recommendation provides a public signal that, if believed, induces individuals to update their expectations about how many others have abandoned the norm or are likely to do so in the current period. By contrast, leadership does not influence the individuals' beliefs in self-efficacy (γ_i), which reflect more stable individual predispositions or preferences, including their evolution over time.

More formally, in the presence of a leader who can send a signal $s_L \in \{0, 1\}$, where $s_L = 0$ if the leader recommends Blue and $s_L = 1$ if they recommend Green, the probability that an individual will choose Green in period t becomes:

$$\mathbb{P}(\text{choice}_t = \text{Green}) = \mathbb{P}(f_i \leq f_{\text{Green},t-1} + \lambda_i (s_L - f_{\text{Green},t-1})) \quad [9]$$

The parameter $\lambda_i \in [0, 1]$ captures how credible an individual deems the leader's recommendation, or equivalently, how much they update their expectations about others' behavior based on the leader's signal. This allows for the possibility that the leader's recommendation is discounted or ignored ($\lambda_i = 0$) when it is perceived to embed a conflict of interest (for example, when the leader's preferences diverge from those of the majority). Note that λ_i may reflect a general trait, such as how much an individual generally trusts a leader, but it may also depend on the leader's past actions. For example, a leader who frequently alternates their recommendations may be deemed less credible over time. How λ_i originates or evolves is not central to our analysis and is therefore not modeled.

From Eq. (9), it follows that individuals are more (less) likely to choose Green when the leader recommends Green (Blue), provided that this recommendation is deemed credible. We can therefore characterize the actual proportion of individuals choosing Green in the previous period that is required for tipping to occur in period t , relative to the baseline condition without a leader, as a function of the leader's recommendation and its credibility. This is illustrated in Fig. S1: the more credible the leader's recommendation for Green (Blue), the smaller (larger) the critical mass required for tipping, and thus the more (less) likely the society is to tip.

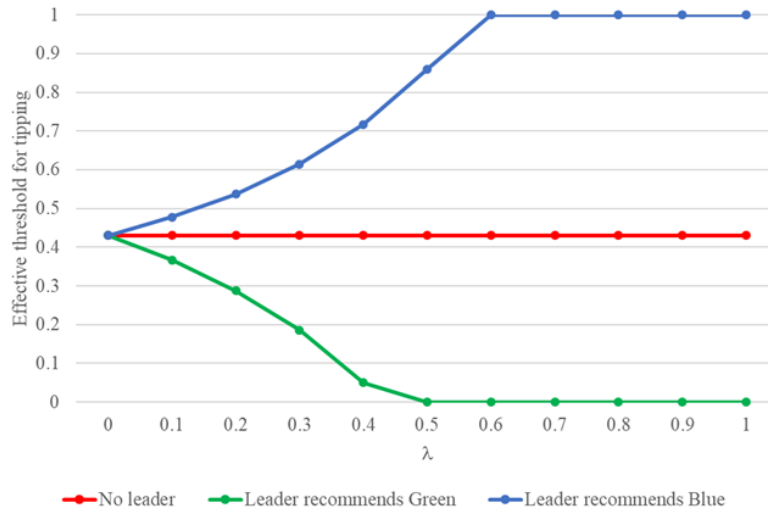


Fig. S1. Effective critical mass for tipping given a leader signal and its perceived credibility (λ). The figure shows how a leader's recommendation shifts the critical mass condition for tipping through belief updating. In the baseline condition without a leader, tipping occurs if the proportion of individuals choosing Green at the beginning of period $t - 1$ exceeds the tipping threshold f_T (i.e., if $f_{\text{Green},t-1} \geq f_T = 0.43$). This threshold is represented by the red line. Under leadership, the tipping condition becomes $\tilde{f}_{\text{Green},t-1} > f_T$, where $\tilde{f}_{\text{Green},t-1} = f_{\text{Green},t-1} + \lambda_i (s_L - f_{\text{Green},t-1})$ denotes the perceived proportion of individuals choosing Green after incorporating the leader's signal s_L . Solving for $f_{\text{Green},t-1}$, this yields the effective critical mass condition $f_{\text{Green},t-1} \geq \frac{f_T - \lambda s_L}{1 - \lambda}$. If the leader recommends Green ($s_L = 1$), the effective critical mass is $\frac{f_T - \lambda}{1 - \lambda}$, which is represented by the green line. If the leader recommends Blue ($s_L = 0$), the effective mass is $f_{\text{Green},t-1} \geq \frac{f_T}{1 - \lambda}$, which is represented by the blue line.

C. Additional data analyses.

C.1. Norm abandonment across treatments. Table S2 displays the average fraction of citizens choosing Green by treatment and time intervals of five periods. It shows that without a leader, Slow-evolving societies were never able to abandon the Blue norm, even when the majority of the citizens had switched to preferring Green, and a change of norms occurred in approximately half of the Fast-evolving societies, but the evolution was slow. In both types of societies, the introduction of a leader made a dramatic change (particularly in Slow-evolving societies) in the fraction of citizens choosing Green.

Table S3 reports the coefficients from a random effects OLS model, along with standard errors clustered at the session level. We regressed the choice of Green on treatment dummies, using the Slow-evolving societies without a leader as the reference category. Model (1) is tested on the data from the last 10 periods, while model (2) is tested on the data from the last five periods. Note that in one Slow-evolving society with a Fast-evolving leader, the game was interrupted after period 14 due to a technical problem. Consequently, the data for the last five periods of this society correspond to periods 10 through 14 inclusive, while the data for the last ten periods correspond to periods 5 through 14 inclusive. Excluding this society from the analysis does not alter our results.

The regressions show that the probability of choosing Green is significantly higher in any treatment compared to the Slow-evolving Baseline, regardless of the time window. We also conducted tests to compare the regression coefficients within each model, yielding results consistent with the non-parametric tests reported in the main text. These parametric tests confirm that citizens in Fast-evolving societies are more likely to choose Green than those in Slow-evolving societies ($p < 0.001$), and highlight the significant impact of the leader in both Slow-evolving ($p < 0.001$) and Fast-evolving societies ($p = 0.011$). Finally, a difference-in-difference analysis comparing the effect of the presence of a leader in each type of society revealed that a Fast-evolving leader in a Slow-evolving society made a bigger difference in the probability of citizens choosing Green compared to an equivalent society without a leader than a Slow-evolving leader in a Fast-evolving society. In model (1), the diff-in-diff coefficient is equal to 0.554 ($p < 0.001$), and in model (2) it is equal to 0.591 ($p < 0.001$).

Similar regressions, including controls for individual characteristics, are available in Table S4. Risk attitude is the only socio-demographics characteristic that is significant at the 5% level.

C.2. Leadership style and individual characteristics. In Table S5, we explore whether the adoption of a given leadership style (democratic *vs.* autocratic *vs.* compromise) was associated with particular individual characteristics of the leaders, including gender, age, level of education, school, risk attitude, trust, patience, compliance with rules, political orientation, and non-conformism. The only notable difference, which is statistically significant in both multivariate and bivariate tests, pertained to the leaders' reported risk attitude. Specifically, democratic leadership was associated with a higher level of risk aversion among leaders compared to autocratic and compromise leadership styles (respective scores of self-reported risk-seeking: 4.5 *vs.* 7.3 *vs.* 7.3 on a 0-10 scale; Kruskal-Wallis test: $p = 0.068$; Mann-Whitney tests: $p = 0.037$ between autocratic and democratic styles, and $p = 0.061$ between democratic and compromise styles). We lack sufficient statistical power to detect other significant differences.

C.3. Citizens as Instigators of Change. In the article, we examine the impact of individuals designated as leaders on norm dynamics. However, it is also interesting to explore the role of citizens who actively sought to initiate a change on their own. We define a citizen as an instigator of change if they chose Green either before or at the same time as the majority of society. This definition excludes individuals who were reluctant to switch to Green before knowing that the majority had established a new norm. Table S11 summarizes the mean characteristics of these citizens in terms of their own preference-switching period, the first period in which they chose Green, and whether they were Fast-evolving citizens. Table S12 replicates Table S11 using a stricter definition of instigators of change, including only those citizens who chose Green before the majority did so. This exclusion considerably reduces the number of instigators of change in the treatments with a leader, as the majority of citizens followed the leader as soon as the latter recommended Green. In the remainder of the analysis, we use the less stringent definition, as we posit that the critical factor is the initiation of the choice of Green when it was still a minority behavior in the previous periods.

According to this definition, out of 800 citizens (20×4 treatments), 56% tried to instigate a change at least once (448/800). In the absence of a leader, there were more instigators of change in Fast-evolving societies than in Slow-evolving societies (110 or 55% *vs.* 40 or 20%). This may not be surprising, as individuals in the former societies should anticipate that the preferences of the majority shift quickly, making an early choice of Green more hopeful compared to a Slow-evolving society. Moreover, more people should be willing to expedite a change, anticipating that a larger proportion of others would prefer Green and be likely to follow suit. In Slow-evolving societies without a leader, 30% of the instigators of change (12/40) were part of the Fast-evolving minority. The percentage of instigators of change in Slow-evolving societies was too low to form a critical mass capable of steering a society, predominantly composed of individuals with slowly evolving preferences, toward a new norm.

In the leader treatments, there were more instigators of change than in the baseline treatments: 168 (84% of the citizens compared to 55% in the corresponding baseline) in Fast-evolving societies and 130 in Slow-evolving societies (65% compared to 20% in the corresponding baseline). The presence of a leader encouraged the instigators of change in both types of society. The increase was particularly prominent in Slow-evolving societies.

For the Baseline treatments, Tables S13 and S14 show that risk attitude differs significantly (at the 5% level) between instigators of change and other members. In contrast, when we pool the data from all treatments (Tables S15 and S16), risk attitude is no longer significant. This contrast suggests that the motivations of early movers are context-dependent and vary with the presence or absence of a coordination device in the society, which can attenuate or substitute for individual risk-taking in driving norm change. Tables S13 and S14 also show a significant gender effect but no other differences in terms of socio-demographic characteristics.

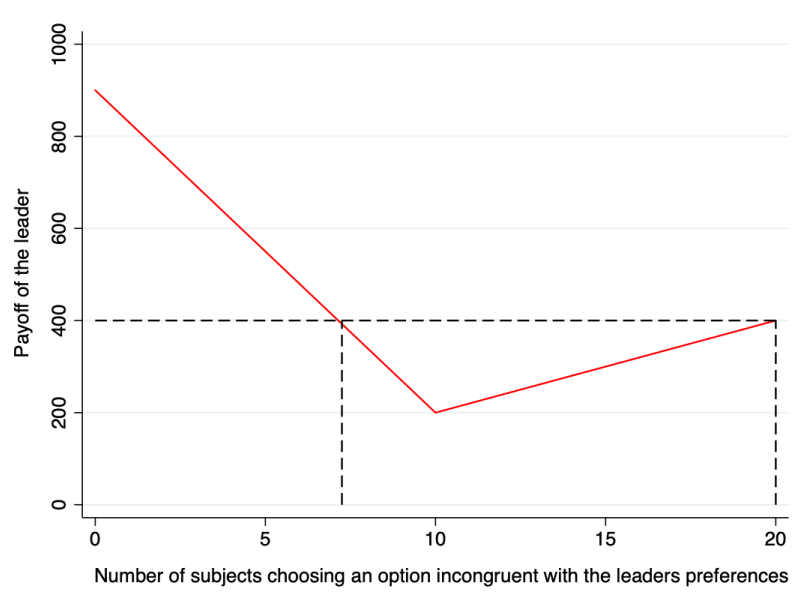


Fig. S2. Payoff function of the leader. In the leader treatments, the leader's payoff is higher if all citizens choose the leader's preferred color than if they all choose the alternative color (900 vs. 400 ECU). The leader's payoff function is V-shaped to express the leader's dislike of normative conflict in society; achieving coordination on a single color, even if not the preferred one, is better for the leader than facing a fully divided society (e.g., half the citizens choosing Blue and the other half choosing Green).

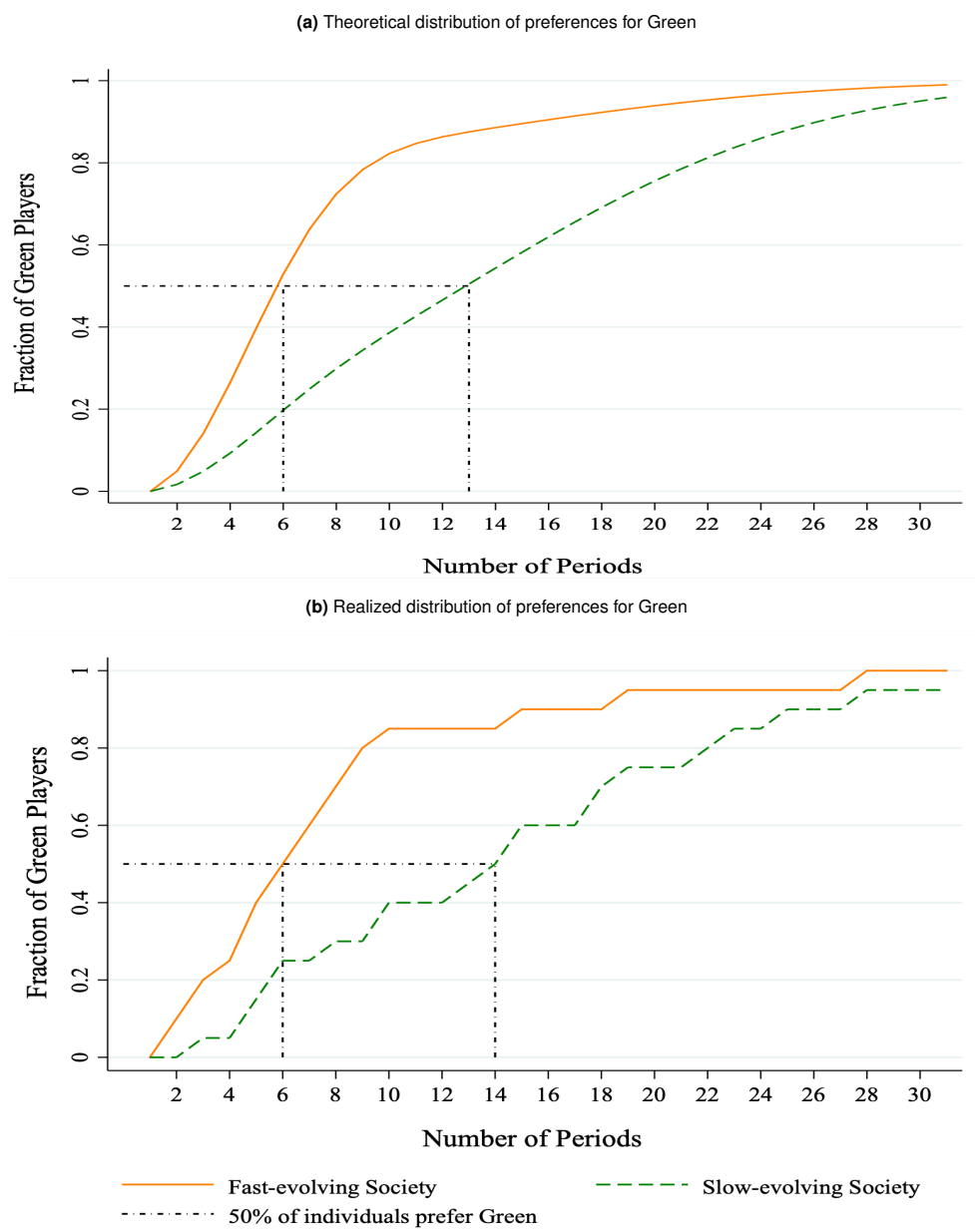


Fig. S3. Theoretical and realized distributions of preferences over time. In “Fast-evolving societies” (solid lines), 80% of citizens changed their preferences from Blue to Green, with a probability that increased by 6 percentage points in each subsequent period (starting from period 2). In contrast, the remaining 20% experienced an increase of only 0.6 percentage points per period, representing a “Slow-evolving minority”. In “Slow-evolving societies” (dashed lines), the probability of 80% of citizens changing preferences increased cumulatively by 0.6 percentage points per period. The probability for the remaining 20% increased cumulatively by 6 percentage points per period, representing a “Fast-evolving minority”. To keep the dynamics of preference changes constant across sessions within the same type of society, we predetermined when each subject would change preference and applied the same distribution of preference changes in each session of a given condition. We did so by generating, for each type of society, 1,000 draws that followed the theoretical distribution of probabilities for that society (represented in Panel (a)). We then selected the distribution (represented in Panel (b)) that minimized the sum of squared errors between the expected and actual distributions. According to the selected distribution, the majority of citizens preferred Green in periods 7 and 15 in the Fast-evolving and Slow-evolving societies, respectively. Thus, for the majority, it became beneficial to coordinate on the Green norm from period 7 in Fast-evolving societies and from period 15 in Slow-evolving societies. The vertical dotted lines indicate the periods (period 6 for Fast-evolving societies and period 14 for Slow-evolving societies) in which preferences were evenly split between Blue and Green. By the end of the game, 100% of subjects prefer Green in Fast-evolving societies and 95% in Slow-evolving societies.

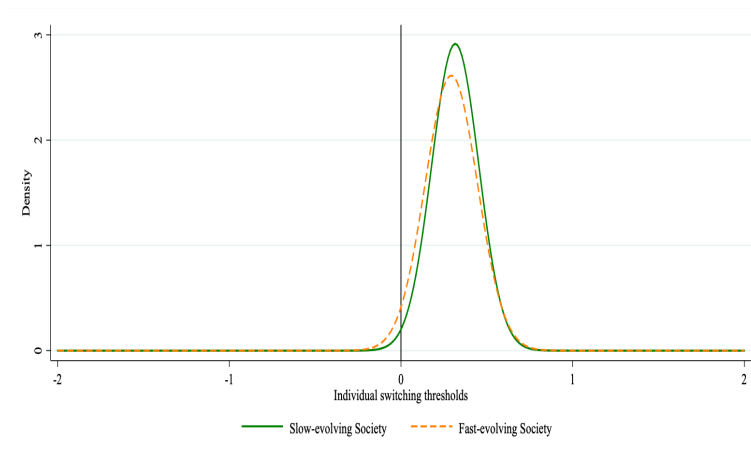


Fig. S4. Estimated distributions of the individual thresholds in Slow-evolving and Fast-evolving societies (Baseline conditions). In both the Slow- and Fast-evolving societies in the baseline conditions, almost all citizens had a positive switching threshold, indicating that they were reluctant to instigate change. In FS-baseline societies (dashed orange line), the distribution of individual tipping thresholds is slightly shifted to the left compared to that in SS-baseline societies with a few more individuals having a threshold equal to or lower than zero, indicating a willingness to initiate change independently of the number of other people who have chosen the Green norm.

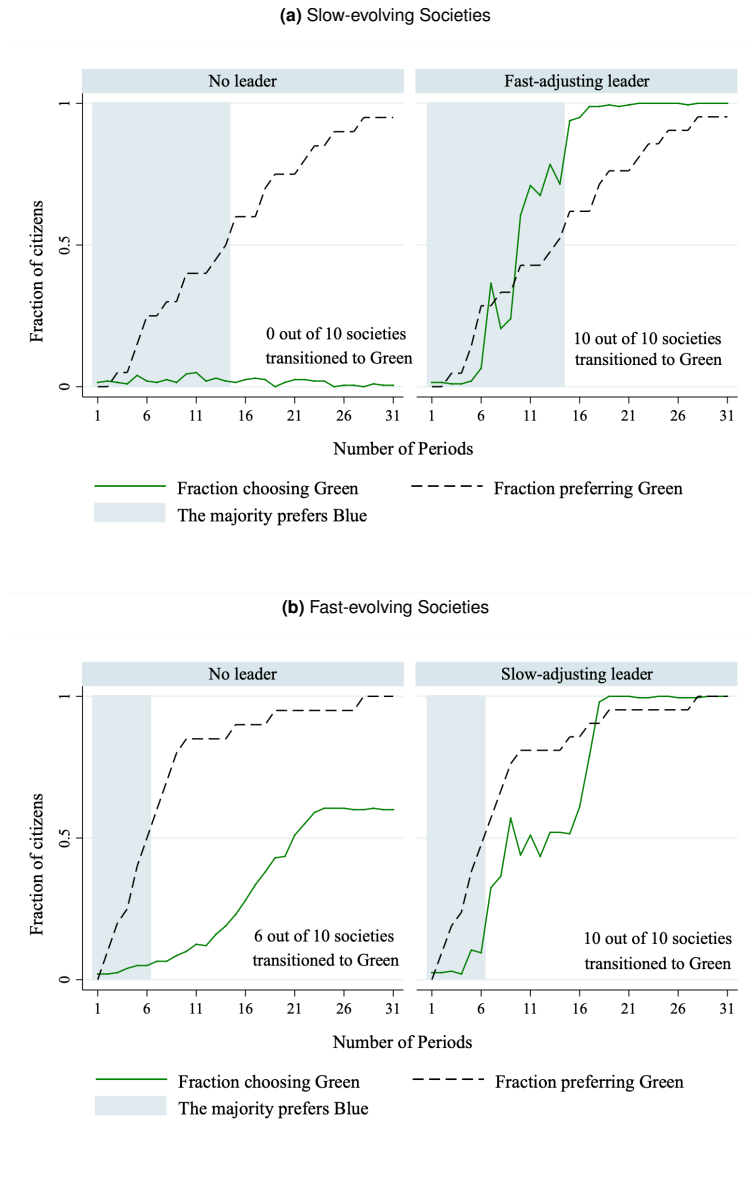


Fig. S5. Preferences for the Green norm (dashed lines) and observed choices (solid lines) of Green in Slow-evolving (Panels (a)) and Fast-evolving (Panels (b)) societies, in the absence (left panels) and presence (right panels) of a leader. The blue-shaded area indicates periods in which the majority of society preferred the Blue norm, while the white area indicates periods in which the majority preferred the Green norm. All leaders preferred the Green norm from period 6 if they were Fast-evolving, and from period 17 if they were Slow-evolving. A society is classified as having transitioned if a majority of individuals coordinate on the Green norm during the final five periods of the game. To observe the dynamics of each society see *SI Appendix, Fig. S6*.

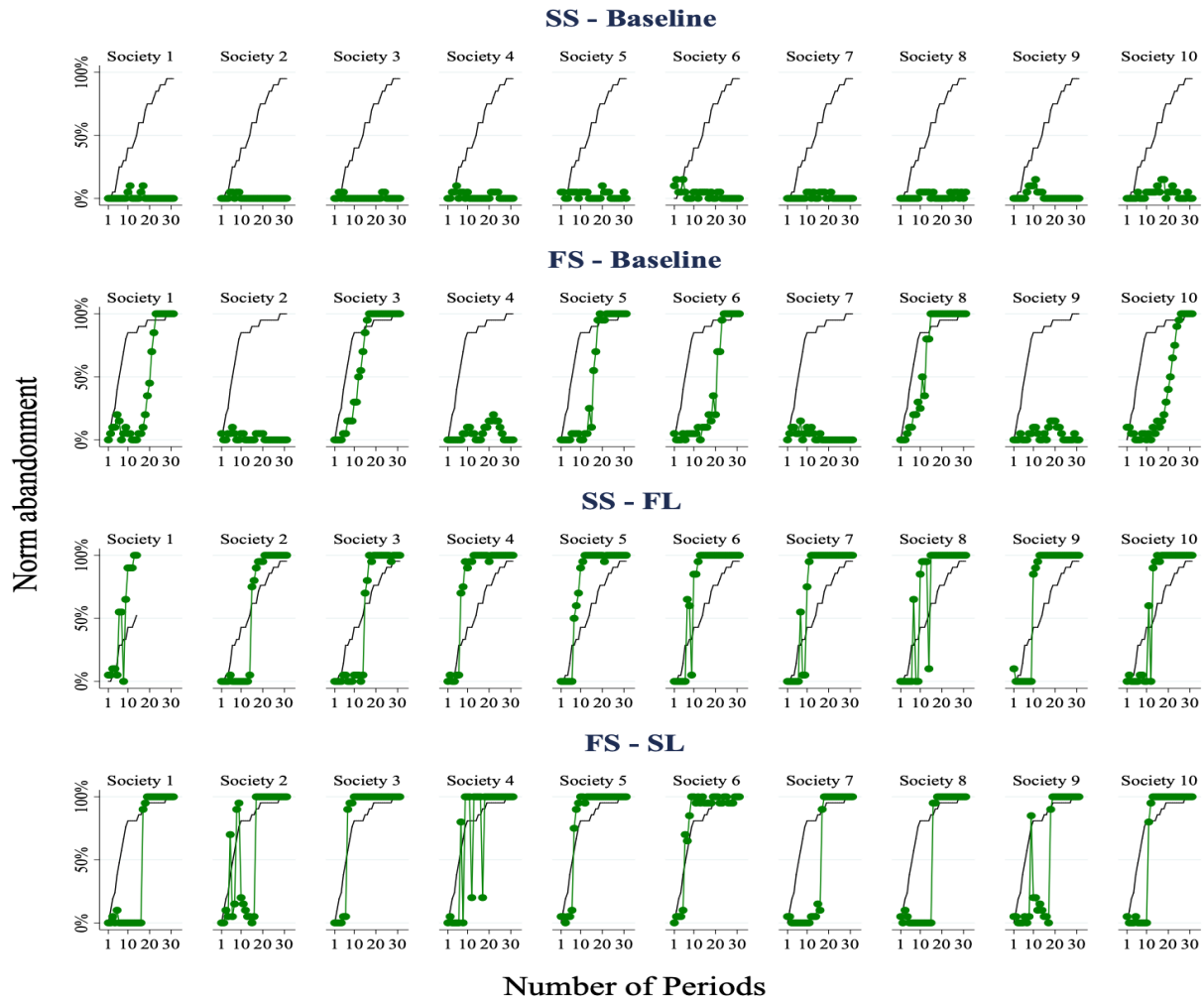
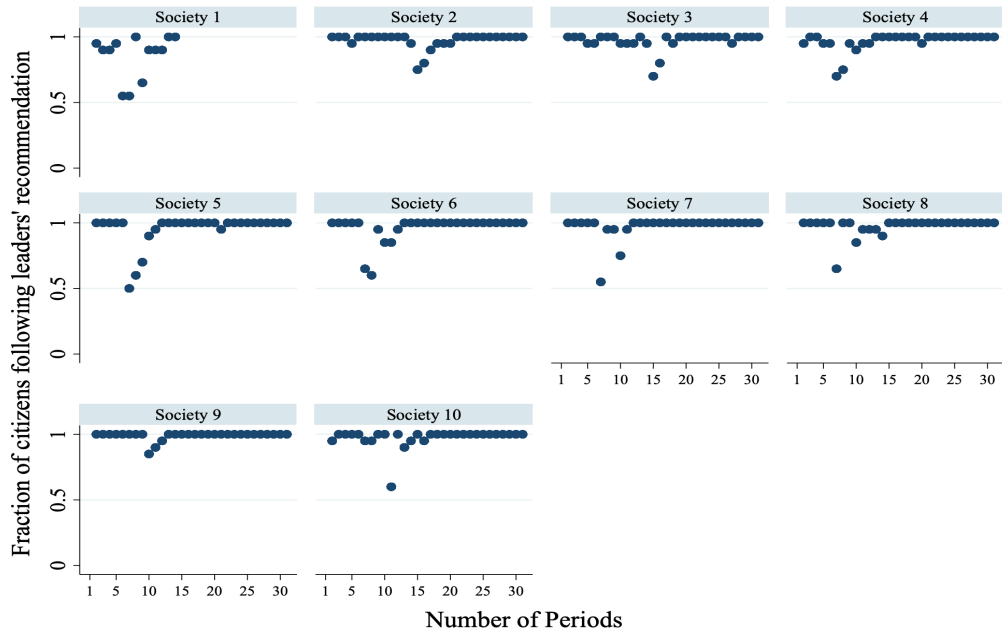


Fig. S6. Norm dynamics in each society, by treatment and type of society. The green line with circled markers represents the citizens' choices of Green and, therefore, indicates norm abandonment. The solid black line shows the fraction of subjects who preferred to abandon the norm. No Slow-evolving society (SS-Baseline) abandoned the initially established norm. Only 60% of the Fast-evolving society (FS-Baseline) abandoned the norm. Both Fast-evolving societies with a Slow-evolving leader (FS-SL) and Slow-evolving societies with a Fast-evolving leader (SS-FL) were able to switch to a better norm. In society 1 of the SS-FL treatment, the series is not complete because the software crashed in period 15.

(a) Slow-evolving Society - Fast-evolving Leader



(b) Fast-evolving Society - Slow-evolving Leader

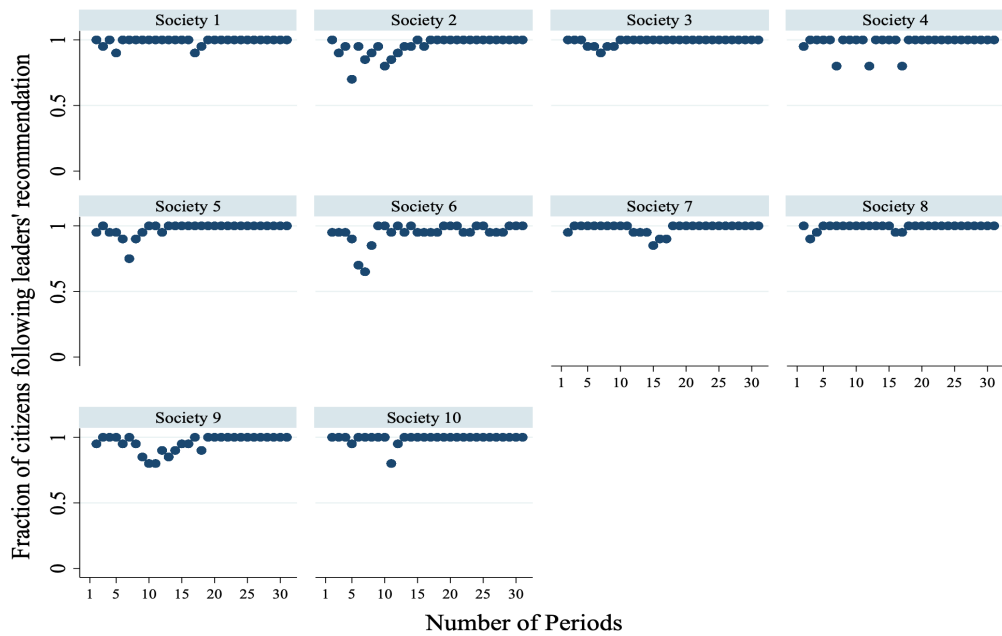


Fig. S7. Fraction of citizens following the leader's recommendation. In the "Slow-evolving Society - Fast-evolving Leader" treatment, the series is incomplete in society 1 because the software crashed in period 15.

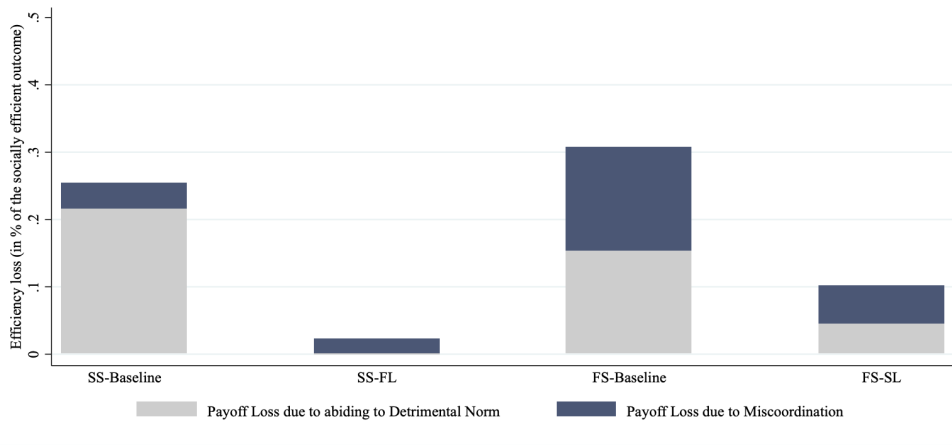


Fig. S8. Citizens' average foregone payoffs when the majority prefers Green. In Fast-evolving (Slow-evolving) societies, the majority preferred Green from period 7 (15). The average foregone payoffs are expressed in percentage relative to the socially efficient outcome, which is 29 ECUs in a Fast-evolving society and 28.08 ECUs in a Slow-evolving society. The figure differentiates between the efficiency loss attributed to adhering to the detrimental norm for too long and the loss resulting from miscoordination penalties. SS-Baseline and FS-Baseline refer to the Slow-evolving and Fast-evolving societies without a leader, respectively. SS-FL refers to Slow-evolving societies with a Fast-evolving leader, and FS-SL refers to Fast-evolving societies with a Slow-evolving leader. In the absence of a leader, societies forfeited approximately one-fourth of what they could have earned by choosing the color associated with the socially efficient outcome. Fast-evolving societies suffered as much as Slow-evolving societies from remaining stuck in the detrimental norm ($p = 0.448$, Mann–Whitney test), but significantly more from penalties incurred when the two subjects in a pair failed to coordinate their actions ($p = 0.001$, Mann–Whitney test). Introducing a leader propelled both types of society toward the socially efficient outcome. The presence of a leader significantly reduced the foregone payoffs due to adherence to the detrimental social norm in both types of society ($p < 0.001$ for Slow-evolving societies and $p = 0.010$ for Fast-evolving societies; Mann–Whitney tests). It also significantly reduced the cost of miscoordination in Fast-evolving societies ($p = 0.002$, Mann–Whitney test), but not in Slow-evolving societies ($p = 0.157$, Mann–Whitney test).

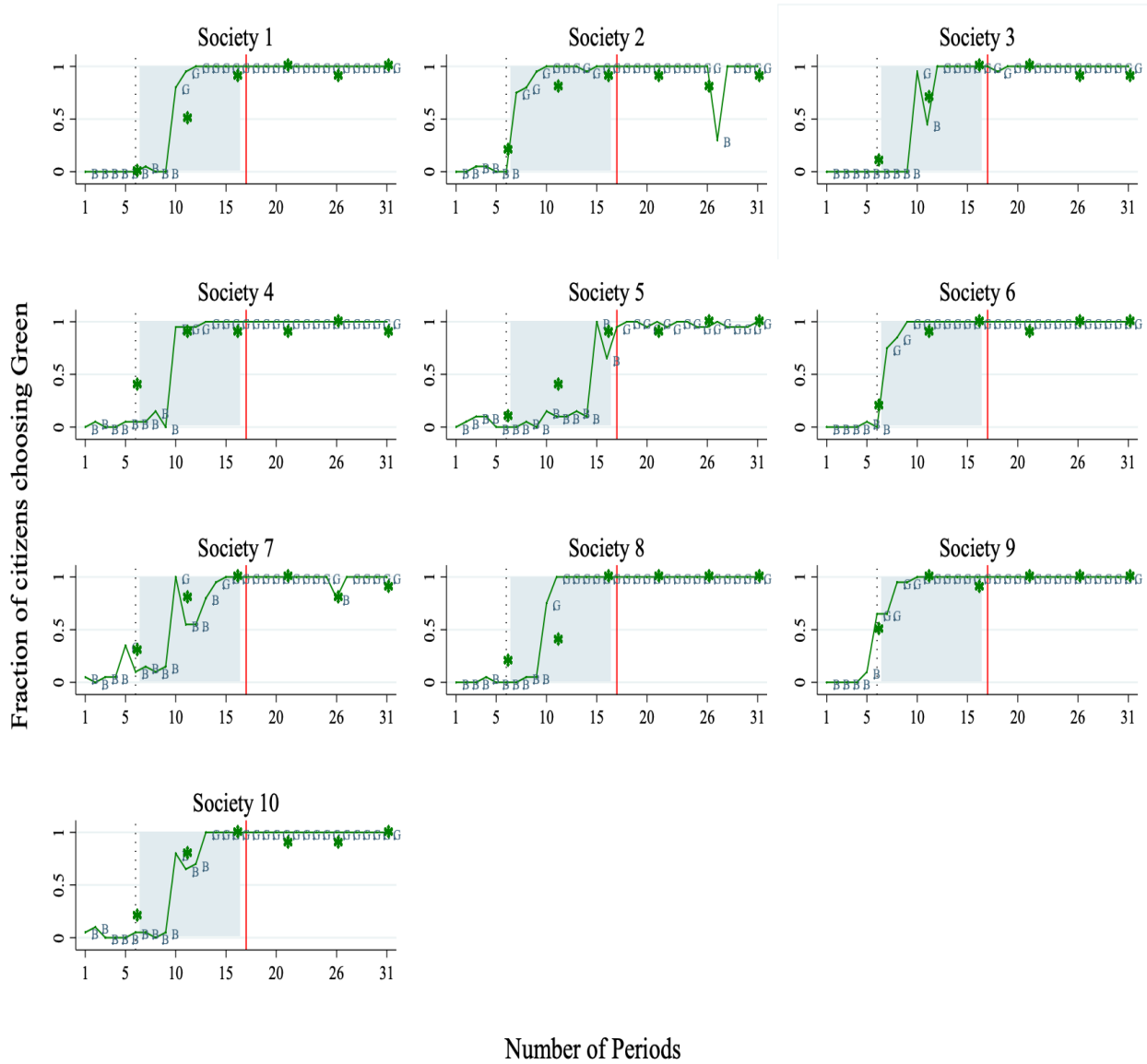


Fig. S9. Evolution of leaders' and citizens' recommendations and the proportion of citizens choosing Green in the FS-SL-Comm treatment. The green solid line represents the fraction of citizens choosing Green over 31 periods in each society of the FS-SL-Comm treatment. The capital letters "B" (for Blue) and "G" (for Green) indicate the recommendation given by the leader in each period. The asterisk (*) indicates the fraction of the 10 messages shown to citizens that recommended Green during periods when peer-to-peer communication was allowed. The vertical line marks the period in which the leader's preference changed from Blue to Green. The blue-shaded area highlights the periods in which the leader's preference and those of the majority of citizens were misaligned.

Table S2. Average fraction of citizens choosing Green, by treatment and time intervals

	Periods					
	1- 5	6-10	11- 15	16-20	21-25	26-31
SS-baseline	0.02 (0.03)	0.02 (0.02)	0.03 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.01)
SS-FL	0.01 (0.02)	0.30 (0.24)	0.76 (0.33)	0.98 (0.03)	1.00 (0.003)	1.00 (0.003)
<i>p</i> -value	0.90	0.02	p<0.01	p<0.01	p<0.01	p<0.01
FS-Baseline	0.03 (0.03)	0.07 (0.06)	0.17 (0.25)	0.37 (0.40)	0.57 (0.46)	0.60 (0.51)
FS-SL	0.04 (0.05)	0.36 (0.35)	0.50 (0.48)	0.88 (0.14)	1.00 (0.01)	1.00 (0.01)
<i>p</i> -value	0.62	0.44	0.15	0.02	0.002	0.09
Total	0.03 (0.03)	0.19 (0.25)	0.36 (0.42)	0.55 (0.44)	0.64 (0.47)	0.64 (0.48)
<i>N</i>	200	200	199	195	195	234

Notes: This table reports the fraction of citizens in the society that chose Green, averaged across societies, with standard deviations in parentheses, by blocks of five periods (except for the last block, where we included the last six periods). The values colored in green refer to the situation where the majority of the citizens preferred Green. The *p*-values correspond to a Mann-Whitney test with each society taken as an independent observation. *N* = 200 corresponds to 10 observations × 5 periods × 4 treatments. The differences in the number of observations across blocks of periods are due to the last 17 periods missing in the session of the SS-FL treatment, where the software crashed.

Table S3. Determinants of the choice of Green by the citizens

Dep. var.: Choice of Green	Last 10 periods (1)	Last 5 periods (2)
Slow-evolving society-Fast-evolving leader	0.955*** (0.034)	0.988*** (0.006)
Fast-evolving society	0.586*** (0.151)	0.596*** (0.157)
Fast-evolving society-Slow-evolving leader	0.988*** (0.004)	0.993*** (0.004)
Constant	0.010*** (0.003)	0.005* (0.003)
<i>N</i>	8000	4000
R ²	0.688	0.722

Notes: This table presents the estimates from random effects linear probability models with standard errors clustered at the session level in parentheses. The dependent variable is a binary variable for the choice of Green in the last 10 periods (model (1)) or the last five periods (model (2)). The omitted category is the Slow-evolving society Baseline. The influence of a Fast-evolving leader in a Slow-evolving society is captured by the first coefficient of the regression, which is 0.955 in model (1) and 0.988 in model (2). The impact of a Slow-evolving leader in a Fast-evolving society is captured by the difference between the third and second coefficients in the regression, which is 0.988 – 0.586 in model (1) and 0.993 – 0.596 in model (2). The diff-in-diff coefficient is obtained by computing the difference between the effect of having a Fast-evolving leader in a Slow-evolving society and the effect of having a Slow-evolving leader in a Fast-evolving society. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table S4. Determinants of the choice of Green by the citizens, with controls for socio-demographics

Dep. var.: Choice of Green	Last 10 periods (1)	Last 5 periods (2)
Slow-evolving soc.-Fast-evolving leader	0.987*** (0.008)	0.990*** (0.008)
Fast-evolving society	0.585*** (0.149)	0.595*** (0.155)
Fast-evolving soc.-Slow-evolving leader	0.989*** (0.010)	0.993*** (0.010)
Risk attitude	0.009** (0.004)	0.010** (0.004)
Patience	0.001 (0.003)	0.000 (0.003)
Trust	0.000 (0.004)	0.000 (0.004)
Political orientation	-0.002 (0.005)	-0.002 (0.005)
Level of education	0.002 (0.014)	0.001 (0.015)
Age	-0.001 (0.003)	-0.001 (0.003)
Rule-following	0.000 (0.001)	0.000 (0.001)
Non-conformist	0.022 (0.028)	0.018 (0.029)
Male	0.004 (0.022)	0.007 (0.023)
Business school	0.030 (0.039)	0.024 (0.040)
Constant	-0.055 (0.079)	-0.048 (0.079)
<i>N</i>	7800	3900
<i>R</i> ²	0.722	0.727

Notes: This table presents the estimates from random effects linear probability models with standard errors clustered at the session level (in parentheses). The dependent variable is a binary variable for the choice of Green in the last 10 periods (model (1)) or the last 5 periods (model (2)). The omitted category is the Slow-evolving society Baseline. Risk attitudes, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). For building the measure of non-conformism, we presented subjects with seven adjective pairs, and they had to indicate on a scale from 1 to 5 the extent to which one adjective in each pair described them better or worse than the other one. After adding the answers to the seven pairs of adjectives, a subject was defined as “Non-conformist” if they scored below the median, and 0 otherwise. 200 observations are missing in model (1) and 100 in model (2) because z-Tree crashed in one session (society 1 in the SS-FL treatment). *** p<0.01, * p<0.1

Table S5. Characteristics of leaders, by leadership style

	Democratic (1)	Autocratic (2)	Compromise (3)	<i>p</i> -value (4)
Risk attitude	4.50 (2.35)	7.30 (2.21)	7.33 (1.15)	0.07*
Patience	6.50 (2.07)	6.10 (2.28)	5.33 (2.89)	0.83
Trust	5.00 (2.28)	5.50 (2.01)	6.67 (1.53)	0.53
Political orientation	4.33 (1.03)	4.30 (2.16)	4.00 (2.65)	0.94
Level of education	2.00 (1.41)	2.30 (0.48)	1.67 (0.58)	0.39
Age	22.50 (1.38)	22.00 (1.49)	21.33 (0.58)	0.39
Rule-following	31.50 (18.59)	14.70 (17.47)	29.00 (19.31)	0.07*
Non-conformist	0.33 (0.52)	0.30 (0.48)	0.67 (0.58)	0.502
Male	0.17 (0.41)	0.30 (0.48)	0.67 (0.58)	0.31
Business school	0.17 (0.41)	0.40 (0.52)	0 (0)	0.31
<i>N</i>	6	10	3	

Notes: This table summarizes the average socio-demographic characteristics of the leaders adopting a top-down leadership style (societies 4, 5, 6, 7, and 8 in the SS-FL treatment and 1, 2, 7, 8, and 9 in the FS-SL treatment) and those of the leaders adopting a bottom-up leadership style (societies 2 and 3 in the SS-FL treatment, and societies 3, 4, 5, and 6 in the FS-SL treatment). The third column shows the characteristics of the leaders who seek compromise (societies 10 and 9 in the SS-FL treatment, and 10 in the FS-SL treatment). Standard deviations are in parentheses. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Kruskal–Wallis *t*-tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. One leader is missing in the top-down leadership category because the software crashed in one session (Slow-evolving Society 1). * $p < 0.10$.

Table S6. Citizens' and leaders' average earnings, by leadership style

	Citizens' payoffs			Leader's payoffs		
	SS-FL	FS-SL	All	SS-FL	FS-SL	All
Autocratic	785.2 (15.56)	776.3 (32.52)	780.75 (24.49)	819.94 (12.87)	849.61 (38.27)	834.77 (31.13)
Democratic	805.5 (9.89)	823 (43.23)	817.17 (34.97)	721.29 (1.37)	717.98 (24.88)	719.09 (19.36)
<i>p</i> -value	0.121	0.086	0.039	0.053	0.014	0.001
<i>N</i>	140	180	320	7	9	16

Notes: This table summarizes the average payoffs of the citizens (left panel) and the leaders (right panel) in the 31 periods, by treatment, depending on the style of leadership adopted by the leader. We restricted the analysis to societies whose leaders could be classified as autocratic or democratic. Standard deviations are in parentheses. The reported *p* values are from two-tailed Mann-Whitney tests, taking each society as an independent observation. SS-FL is for Slow-evolving societies with a Fast-evolving leader, while FS-SL is for Fast-evolving societies with a Slow-evolving leader. 20 observations are missing for the citizens' payoffs and one for the leader's payoffs because the program crashed in one session (Slow-evolving society - Fast-evolving leader 1).

Table S7. Socio-demographic characteristics of the citizens, by treatment

	SS-Base (1)	SS-FL (2)	FS-Base (3)	FS-SL (4)	<i>p</i> -value (5)
Risk attitude	5.81 (2.20)	6.16 (2.25)	6.00 (2.20)	6.09 (2.32)	0.277
Patience	5.68 (2.48)	5.62 (2.65)	5.54 (2.63)	5.83 (2.57)	0.780
Trust	4.99 (2.40)	5.14 (2.36)	4.79 (2.31)	5.15 (2.33)	0.390
Political orientation	4.59 (2.28)	4.43 (2.00)	4.46 (2.08)	4.57 (2.09)	0.832
Level of education	2.21 (1.10)	2.28 (1.16)	2.29 (1.20)	2.21 (1.10)	0.895
Age	22.61 (1.94)	23.01 (3.09)	23.59 (2.97)	22.43 (2.08)	0.001***
Rule-following	18.20 (20.60)	21.63 (18.78)	16.70 (19.70)	21.55 (20.22)	0.003***
Non-conformist	0.47 (0.50)	0.51 (0.50)	0.50 (0.50)	0.49 (0.50)	0.860
Male	0.49 (0.50)	0.47 (0.50)	0.46 (0.50)	0.57 (0.50)	0.093*
Business school	0.48 (0.50)	0.42 (0.49)	0.48 (0.50)	0.28 (0.45)	<0.001***
<i>N</i>	200	180	200	200	

Notes: This table summarizes the average socio-demographic characteristics of participants acting as citizens per treatment, with standard deviations in parentheses. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Kruskal–Wallis *t*-tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). For building the measure of non-conformism, we presented subjects with seven adjective pairs and they had to indicate on a scale from 1 to 5 the extent to which one adjective in each pair described them better or worse than the other one. After adding the answers to the seven pairs of adjectives, a subject was defined as “Non-conformist” if they scored below the median, and 0 otherwise. The *p*-values reported for these variables are from Chi-squared tests. SS-Base is for Slow-evolving societies without a leader, and SS-FL is for Slow-evolving societies with a Fast-evolving leader; FS-Base is for Fast-evolving societies without a leader, and FS-SL is for Fast-evolving societies with a Slow-evolving leader. 20 observations are missing in the SS-FL treatment because the software crashed during one session (society 1). *** $p < 0.01$, * $p < 0.1$.

Table S8. Socio-demographic characteristics of the leaders, by treatment

	SS-FL (1)	FS-SL (2)	<i>p</i> -value (3)
Risk attitude	7.11 (2.47)	5.80 (2.35)	0.246
Patience	6.33 (2.29)	5.90 (2.23)	0.650
Trust	5.11 (2.52)	5.90 (1.45)	0.618
Political orientation	4.33 (2.55)	4.20 (1.03)	0.868
Level of education	1.89 (0.93)	2.30 (0.82)	0.437
Age	22.22 (1.20)	21.90 (1.52)	0.608
Rule-following	14.33 (21.07)	29.40 (14.19)	0.067*
Non-conformist	0.22 (0.44)	0.50 (0.53)	0.210
Male	0.33 (0.50)	0.30 (0.48)	0.876
Business school	0.33 (0.50)	0.20 (0.42)	0.510
<i>N</i>	9	10	

Notes: This table summarizes the average socio-demographic characteristics of the leaders per treatment, with standard deviations in parentheses. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative stance. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Kruskal–Wallis *t*-tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. SS-FL is for Slow-evolving societies with a Fast-evolving leader, and FS-SL is for Fast-evolving societies with a Slow-evolving leader. One observation is missing from the SS-FL treatment because the software crashed during one session (society 1). ** $p < 0.05$.

Table S9. Socio-demographic characteristics of the citizens, follow-up study

	FS-SL (1)	FS-SL-Comm (2)	p-value (3)
Risk Attitude	6.09 (2.32)	5.99 (2.20)	0.554
Patience	5.83 (2.57)	5.63 (2.50)	0.480
Trust	5.15 (2.33)	4.79 (2.51)	0.131
Political orientation	4.57 (2.09)	4.62 (2.08)	0.851
Level of education	2.21 (1.10)	2.47 (1.07)	0.009***
Age	22.43 (2.08)	22.03 (2.21)	0.013**
Rule-following	21.55 (20.22)	21.34 (19.62)	0.969
Non-conformist	0.38 (0.49)	0.46 (0.50)	0.128
Male	0.57 (0.50)	0.42 (0.49)	0.002***
Business School	0.28 (0.45)	0.42 (0.50)	0.003***
<i>N</i>	200	200	

Notes: This table summarizes the average socio-demographic characteristics of citizens in FS-SL and FS-SL-Comm treatment, with standard deviations in parentheses. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. FS-SL is for Fast-evolving societies with a Slow-evolving leader, and FS-SL-Comm is for the follow-up treatment with peer-to-peer communication. *** $p < 0.01$, * $p < 0.1$.

Table S10. Socio-demographic characteristics of the leaders, follow-up study

	FS-SL (1)	FS-SL-Comm (2)	p-value (3)
Risk attitude	5.80 (2.35)	6.30 (2.26)	0.729
Patience	5.90 (2.23)	4.40 (2.27)	0.137
Trust	5.90 (1.45)	5.70 (3.20)	0.703
Political orientation	4.20 (1.03)	5.30 (1.83)	0.152
Level of education	2.30 (0.82)	2.30 (1.34)	0.875
Age	21.90 (1.52)	22 (2.05)	0.938
Rule-following	29.40 (14.19)	11 (16.39)	0.008***
Non-conformist	0.30 (0.48)	0.50 (0.53)	0.361
Male	0.30 (0.48)	0.60 (0.52)	0.178
Business School	0.20 (0.42)	0.40 (0.52)	0.329
<i>N</i>	10	10	

Notes: This table summarizes the average socio-demographic characteristics of conservative leaders in the PS-CL and PS-CL-Comm treatment, with standard deviations in parentheses. Risk attitudes, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. PS-CL is for Progressive societies with a Conservative leader and PS-CL-Comm is for the follow-up treatment with peer-to-peer communication. *** $p < 0.01$, * $p < 0.1$.

Table S11. Instigators of change

	SS-Baseline (1)	SS-FL (2)	FS-Baseline (3)	FS-SL (4)	Total (5)
Switch	11.57 (7.48)	14.52 (8.42)	6.72 (4.13)	7.57 (5.79)	9.73 (7.28)
First deviation	12.22 (7.46)	9.08 (3.40)	12.78 (5.79)	9.95 (4.77)	10.60 (5.21)
Share of Fast-evolving citizens	0.30 (0.46)	0.24 (0.43)	0.88 (0.32)	0.82 (0.38)	0.62 (0.49)
<i>N</i>	40	130	110	168	448

Notes: This table reports average values with standard deviations in parentheses. Instigators of change are defined as citizens choosing Green either before or at the same time as the majority of society. SS-Baseline and FS-Baseline refer to Slow-evolving and Fast-evolving societies without a leader, respectively. SS-FL refers to Slow-evolving societies with a Fast-evolving leader, and FS-SL refers to Fast-evolving societies with a Slow-evolving leader. “Switch” indicates the average period in which the preferences of the instigators changed from Blue to Green. ‘First deviation’ indicates the first period on average in which instigators of change started choosing Green. ‘Share of Fast-evolving citizens’ is the proportion of instigators of change that come from the minority in Slow-evolving societies or from the majority in Fast-evolving societies.

Table S12. Instigators of change - stricter definition

	SS-Baseline (1)	SS-FL (2)	FS-Baseline (3)	FS-SL (4)	Total (5)
Switch	11.57 (7.48)	11.78 (8.97)	6.90 (4.36)	7.25 (6.53)	8.39 (6.18)
First deviation	12.22 (7.46)	5.89 (3.95)	12.07 (5.85)	5.55 (3.99)	10.94 (6.47)
Share of Fast-evolving citizens	0.30 (0.46)	0.33 (0.50)	0.88 (0.33)	0.85 (0.37)	0.70 (0.46)
<i>N</i>	40	9	91	20	160

Notes: This table reports average values with standard deviations in parentheses. Instigators of change are defined in this table as citizens who chose Green before the majority did so. SS is for Slow-evolving society, FS for Fast-evolving society, FL for Fast-evolving leader, and SL for Slow-evolving leader. 'Switch' indicates the average period in which the preferences of instigators changed from Blue to Green. 'First deviation' indicates the first period on average in which the instigators of change started choosing Green. 'Share of Fast-evolving citizens' is the proportion of instigators of change that come from the minority in Slow-evolving societies or from the majority in Fast-evolving societies.

Table S13. Socio-demographic characteristics of the instigators - Baseline treatments

	Non-Instigators (1)	Instigators (2)	p-value (3)
Risk attitude	5.72 (2.21)	6.20 (2.16)	0.021***
Patience	5.56 (2.58)	5.69 (2.53)	0.615
Trust	4.95 (2.29)	4.79 (2.47)	0.455
Political orientation	4.53 (2.24)	4.53 (2.08)	0.920
Age	22.90 (2.03)	23.45 (3.22)	0.228
Level of education	2.32 (1.10)	2.13 (1.22)	0.204
Rule-following	18.19 (20.90)	16.20 (18.82)	0.398
Non-conformist	0.48 (0.50)	0.49 (0.50)	0.737
Male	0.52 (0.50)	0.41 (0.49)	0.034**
Business School	0.47 (0.50)	0.50 (0.50)	0.587
<i>N</i>	250	150	

Notes: This table summarizes the average socio-demographic characteristics of the instigators and the non-instigators in the two baseline treatments, with standard deviations in parentheses. Instigators of change are defined as citizens choosing Green either before or at the same time as the majority of society. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests.

Table S14. Socio-demographic characteristics of the instigators - stricter definition

	Non-Instigators (1)	Instigators (2)	p-value (3)
Risk attitude	5.76 (2.19)	6.20 (2.19)	0.041***
Patience	5.59 (2.58)	5.64 (2.52)	0.829
Trust	4.93 (2.29)	4.79 (2.49)	0.515
Political orientation	4.52 (2.20)	4.54 (2.15)	0.950
Age	22.91 (2.00)	23.50 (3.39)	0.353
Level of education	2.32 (1.10)	2.11 (1.22)	0.159
Rule-following	17.87 (20.70)	16.57 (19.01)	0.654
Non-conformist	0.48 (0.50)	0.48 (0.50)	0.965
Male	0.52 (0.50)	0.37 (0.49)	0.005***
Business School	0.48 (0.50)	0.50 (0.50)	0.702
<i>N</i>	269	131	

Notes: This table summarizes the average socio-demographic characteristics of the instigators in the baseline (according to the stricter definition that considers instigators of change only those citizens who chose Green before the majority did so) and the non-instigators, with standard deviations in parentheses. Risk attitude, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and for the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests.

Table S15. Socio-demographic characteristics of the instigators - All treatments

	Non-Instigators (1)	Instigators (2)	p-value (3)
Risk attitude	5.96 (2.19)	6.05 (2.28)	0.341
Patience	5.59 (2.62)	5.72 (2.55)	0.585
Trust	4.93 (2.31)	5.08 (2.38)	0.404
Political orientation	4.51 (2.19)	4.52 (2.05)	0.859
Age	22.91 (2.11)	22.91 (2.92)	0.071*
Level of education	2.30 (1.13)	2.20 (1.14)	0.306
Rule-following	17.98 (20.13)	20.63 (19.74)	0.046**
Non-conformist	0.48 (0.50)	0.49 (0.50)	0.775
Male	0.52 (0.50)	0.49 (0.50)	0.348
Business School	0.45 (0.50)	0.39 (0.49)	0.077*
<i>N</i>	343	437	

Notes: This table summarizes the average socio-demographic characteristics of the instigators and the non-instigators across all treatments (FS-Baseline, SS-Baseline, FS-SL and SS-FL), with standard deviations in parentheses. Instigators of change are defined as citizens choosing Green either before or at the same time as the majority of society. Risk attitudes, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. The number of instigators is 437 instead of 448 because we are missing the socio-demographics of session 1 in SS-FL treatment, where the software crashed. *** $p < 0.01$, * $p < 0.1$.

Table S16. Socio-demographic characteristics of the instigators - stricter definition - all treatments

	Non-Instigators (1)	Instigators (2)	p-value (3)
Risk attitude	5.95 (2.26)	6.22 (2.19)	0.182
Patience	5.67 (2.59)	5.64 (2.53)	0.918
Trust	5.05 (2.30)	4.86 (2.54)	0.299
Political orientation	4.49 (2.10)	4.60 (2.16)	0.599
Age	22.73 (2.14)	23.59 (3.82)	0.027**
Level of Education	2.28 (1.11)	2.13 (1.21)	0.250
Rule-following	19.90 (20.10)	17.77 (19.28)	0.145
Non-conformist	0.49 (0.50)	0.50 (0.50)	0.705
Male	0.53 (0.50)	0.39 (0.49)	0.002***
Business School	0.40 (0.49)	0.47 (0.50)	0.115
<i>N</i>	621	159	

Notes: This table summarizes the average socio-demographic characteristics of the instigators (according to the stricter definition that considers instigators of change only those citizens who chose Green before the majority did so) and the non-instigators across all treatments (FS-Baseline, SS-Baseline, FS-SL and SS-FL), with standard deviations in parentheses. Risk attitudes, patience, and trust are measured on a scale from 1 (lowest) to 10 (highest). Political orientation is also measured on a scale from 1 to 10, with 1 indicating a strongly liberal stance and 10 indicating a strongly conservative one. The level of education is equal to 0 for Bac+1, 1 for Bac+2, and so on (Bac is the final high school exam in France). Rule following is measured on a scale of 0 to 50. The *p*-values reported for these variables and for the age variable are from Mann-Whitney tests. The table also includes binary variables: non-conformist (equal to 1 if the person is non-conformist, 0 otherwise), male (equal to 1 if the person is a male, 0 otherwise), and business school (equal to 1 if the student comes from a business school, 0 otherwise). The *p*-values reported for these variables are from Chi-squared tests. The number of instigators is 159 instead of 160 because we are missing the socio-demographics of session 1 in SS-FL treatment, where the software crashed. *** $p < 0.01$, * $p < 0.1$.

D. INSTRUCTIONS.

Experimental Instructions [*Translated from French*]

Slow evolving Societies (SS - Baseline)

GENERAL INSTRUCTIONS

Hello and welcome to this decision-making experiment!

During this experiment, you can earn money. Your earnings from the experiment will largely depend on your decisions and the decisions of other participants in this experiment. Therefore, it is important to read the instructions carefully.

The amount you will earn at the end of this experiment will be the sum of your earnings during the session and a show-up fee of 5 Euros. If at the end of the session, you have made a loss, you will only receive the show-up fee of 5 Euros. Throughout the experiment, we will not speak of EURO, but rather of ECU (Experimental Currency Units). At the end of the session, the total amount of ECU you earned will be converted into EURO according to the following exchange rate:

$$100 \text{ ECU} = 1 \text{ EURO}$$

All your decisions and answers will be anonymous. You are not allowed to communicate with other participants for the entire duration of the experiment, otherwise, you will be excluded from the session and you will not receive any earnings. You are not allowed to use your phone during the entire session. We kindly ask you to switch it off and put it away.

If you have any questions, please raise your hand or press the red button on the side of your desk. Someone from the experimental team, will come and answer your questions immediately.

The experiment consists of two parts. You will now receive the instructions for Part 1. At the end of Part 1, you will receive the instructions for Part 2.

PART 1

Duration and matching

The first part of the experiment consists of 31 periods. All participants in the session form a group of 20 people.

At the beginning of each period, you will be randomly matched to one of the other 19 participants. You will never know the identity of the person with whom you are matched in a given period.

Your task

Your task in each period is to choose between two actions: **BLUE** or **GREEN**. The participant you are matched with in this period also chooses **BLUE** or **GREEN** at the same time as you. Your earnings from a given period will depend partly on your actions and the actions of the other participant with whom you were matched.

There are four possibilities:

- (i) You both choose **BLUE**.
- (ii) You both choose **GREEN**.
- (iii) You choose **BLUE** and the other participant chooses **GREEN**
- (iv) You choose **GREEN** and the other participant chooses **BLUE**

Your Earnings

Your Final Earnings in Part 1 will be the sum of the earnings you made during the 31 periods.

Your earnings in a given period of the experiment depend on two factors: the choices and the types of participants.

There are two possible types in the experiment: **Type A** and **Type B**. We will explain precisely how the types are determined later on. For the time being, it suffices to know that **all participants start off as Type A**.

-A Type A participant receives **30 ECU** when they choose **BLUE** and **20 ECU** when they choose **GREEN**.

-A Type B participant receives **20 ECU** when they choose **BLUE** and **30 ECU** when they choose **GREEN**.

If the other participant chooses the same color as you, the earnings above are the final earnings in a given period. However, every time you and the other participant you were matched with choose **different colors** (i.e., cases iii and iv above) you both receive a “**miscoordination penalty**”. The penalty may differ for the two participants.

In other words, the amount you will receive (see above) will be reduced by **4 ECU for each participant in the group of 20 participants that chose a different color than you**. That is, the more people choose a different action than you, the greater will be your miscoordination penalty.

You will be informed about the number of participants who chose a different color than you with a delay of one period. Thus, if you and the player who is matched to you chose a different color, you will learn your exact payoff with a delay of one period.

Example 1: Suppose that you and the other participant you were matched with are both Type A. You choose **BLUE** and the other participant chooses **GREEN**. All the other participants in your group choose **BLUE**, except the participant you are matched with who chose **GREEN**. Then you will earn 30 ECU minus (1 participant * 4 ECU) = 26 ECU. The person choosing **GREEN** will earn 20 ECU minus (19 participants * 4 ECU) = - 56 ECU, that is, this participant will lose 56 ECU.

Example 2: Suppose that you are Type B and choose **GREEN**. Assume further, that the participant you were matched with is Type A and chooses **BLUE**. Among the 20 participants, you and another participant in your group are the only ones choosing **GREEN**; the rest choose **BLUE**. Then you will earn 30 ECU minus (18 participants * 4 ECU) = -42 ECU. The other participant will earn 30 ECU minus (2 participants * 4 ECU) = 22 ECU.

Example 3: Suppose that you and the other participant are Type B. You choose **BLUE** and the participant you were matched with chooses **GREEN**. Assume further that among all the 20 participants, 10 choose **GREEN** and 10 choose **BLUE**. Then you will earn 20 ECU minus (10 participants * 4 ECU) = -20 ECU. The person choosing **GREEN** will earn 30 ECU minus (10 participants * 4 ECU) = - 10 ECU.

Type determination and implications for earnings

As mentioned above, **all participants start the experiment as Type A**. The probability of changing type in period 1 is equal to zero. From period 2 onward, each participant has a probability p of becoming Type B. In each period, this probability **increases** with a fixed percentage p .

For 16 participants, the probability p of becoming Type B is 0.6% in period 2 (i.e., they have 0.6 chances out of 100 of becoming Type B in period 2). This probability **increases** by 0.6 percentage points in each subsequent period. This means that their probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on. For the 4 other participants, the probability p of becoming Type B is 6% in period 2 (i.e., they have 6 chances in 100 of becoming Type B in period 2). This probability **increases** by 6 percentage points in each subsequent period. This means that their probability of becoming Type B is 12% in period 3, 18% in period 4, and so on.

These probabilities are randomly assigned and are independent for each participant. You will be informed of your probability when the game starts.

The exact period in which you will change Type has already been determined before the beginning of the session. The selection was done independently for each member of your group, based on simulations that reflect a probability of change that increases with each period by: 0.6 percentage points for 16 participants and by 6 percentage points for 4 participants.

If your type changes to Type B you will be informed of this change at the beginning of the period, but you will not know whether the participant with whom you are matched has changed type or is still Type A. The change in type is definitive: once a participant becomes Type B, they do not change back to Type A.

This completes the description of the first part. Please read these instructions again. If you have any questions, raise your hand or press the red button. Otherwise, you can start answering the training questions on the computer. The purpose of these questions is to make sure you have understood the instructions. Any points that are not clear enough will be re-explained by the experimenter. Once you have answered all the questions correctly, the experiment will start.

Fast evolving Societies (FS - Baseline)

GENERAL INSTRUCTIONS

Hello and welcome to this decision-making experiment!

During this experiment, you can earn money. Your earnings from the experiment will largely depend on your decisions and the decisions of other participants in this experiment. Therefore, it is important to read the instructions carefully.

The amount you will earn at the end of this experiment will be the sum of your earnings during the session and a show-up fee of 5 Euros. If at the end of the session, you have made a loss, you will only receive the show-up fee of 5 Euros. Throughout the experiment, we will not speak of EURO, but rather of ECU (Experimental Currency Units). At the end of the session, the total amount of ECU you earned will be converted into EURO according to the following exchange rate:

$$100 \text{ ECU} = 1 \text{ EURO}$$

All your decisions and answers will be anonymous. You are not allowed to communicate with other participants for the entire duration of the experiment, otherwise, you will be excluded from the session and you will not receive any earnings. You are not allowed to use your phone during the entire session. We kindly ask you to switch it off and put it away.

If you have any questions, please raise your hand or press the red button on the side of your desk. Someone from the experimental team, will come and answer your questions immediately.

The experiment consists of two parts. You will now receive the instructions for Part 1. At the end of Part 1, you will receive the instructions for Part 2.

PART 1

Duration and matching

The first part of the experiment consists of 31 periods. All participants in the session form a group of 20 people.

At the beginning of each period, you will be randomly matched to one of the other 19 participants. You will never know the identity of the person with whom you are matched in a given period.

Your task

Your task in each period is to choose between two actions: **BLUE** or **GREEN**. The participant you are matched with in this period also chooses **BLUE** or **GREEN** at the same time as you. Your earnings from a given period will depend partly on your actions and the actions of the other participant with whom you were matched.

There are four possibilities:

- (i) You both choose **BLUE**.
- (ii) You both choose **GREEN**.
- (iii) You choose **BLUE** and the other participant chooses **GREEN**
- (iv) You choose **GREEN** and the other participant chooses **BLUE**

Your Earnings

Your Final Earnings in Part 1 will be the sum of the earnings you made during the 31 periods.

Your earnings in a given period of the experiment depend on two factors: the choices and the types of participants.

There are two possible types in the experiment: **Type A** and **Type B**. We will explain precisely how the types are determined later on. For the time being, it suffices to know that **all participants start off as Type A**.

-A Type A participant receives **30 ECU** when they choose **BLUE** and **20 ECU** when they choose **GREEN**.

-A Type B participant receives **20 ECU** when they choose **BLUE** and **30 ECU** when they choose **GREEN**.

If the other participant chooses the same color as you, the earnings above are the final earnings in a given period. However, every time you and the other participant you were matched with choose **different colors** (i.e., cases iii and iv above) you both receive a “**miscoordination penalty**”. The penalty may differ for the two participants.

In other words, the amount you will receive (see above) will be reduced by **4 ECU for each participant in the group of 20 participants that chose a different color than you**. That is, the more people choose a different action than you, the greater will be your miscoordination penalty.

You will be informed about the number of participants who chose a different color than you with a delay of one period. Thus, if you and the player who is matched to you chose a different color, you will learn your exact payoff with a delay of one period.

Example 1: Suppose that you and the other participant you were matched with are both Type A. You choose **BLUE** and the other participant chooses **GREEN**. All the other participants in your group choose **BLUE**, except the participant you are matched with who chose **GREEN**. Then you will earn 30 ECU minus (1 participant * 4 ECU) = 26 ECU. The person choosing **GREEN** will earn 20 ECU minus (19 participants * 4 ECU) = - 56 ECU, that is, this participant will lose 56 ECU.

Example 2: Suppose that you are Type B and choose **GREEN**. Assume further, that the participant you were matched with is Type A and chooses **BLUE**. Among the 20 participants, you and another participant in your group are the only ones choosing **GREEN**; the rest choose **BLUE**. Then you will earn 30 ECU minus (18 participants * 4 ECU) = -42 ECU. The other participant will earn 30 ECU minus (2 participants * 4 ECU) = 22 ECU.

Example 3: Suppose that you and the other participant are Type B. You choose **BLUE** and the participant you were matched with chooses **GREEN**. Assume further that among all the 20 participants, 10 choose **GREEN** and 10 choose **BLUE**. Then you will earn 20 ECU minus (10 participants * 4 ECU) = -20 ECU. The person choosing **GREEN** will earn 30 ECU minus (10 participants * 4 ECU) = - 10 ECU.

Type determination and implications for earnings

As mentioned above, **all participants start the experiment as Type A**. The probability of changing type in period 1 is equal to zero. From period 2 onward, each participant has a probability p of becoming Type B. In each period, this probability **increases** with a fixed percentage p .

For 16 participants, the probability p of becoming Type B is 6% in period 2 (i.e., they have 6 chances out of 100 of becoming Type B in period 2). This probability **increases** by 6 percentage points in each subsequent period. This means that their probability of becoming Type B is 12% in period 3, 18% in period 4, and so on. For the 4 other participants, the probability p of becoming Type B is 0.6% in period 2 (i.e., they have 0.6 chances in 100 of becoming Type B in period 2). This probability **increases** by 0.6 percentage points in each subsequent period. This means that their probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on.

These probabilities are randomly assigned and are independent for each participant. You will be informed of your probability when the game starts.

The exact period in which you will change Type has already been determined before the beginning of the session. The selection was done independently for each member of your group, based on simulations that reflect a probability of change that increases with each period by 6 percentage points for 16 participants and by 0.6 percentage points for 4 participants.

If your type changes to Type B, you will be informed of this change at the beginning of the period, but you will not know whether the participant with whom you are matched has changed type or is still Type A. The change in type is definitive: once a participant becomes Type B, they do not change back to Type A.

This completes the description of the first part. Please read these instructions again. If you have any questions, raise your hand or press the red button. Otherwise, you can start answering the training questions on the computer. The purpose of these questions is to make sure you have understood the instructions. Any points that are not clear enough will be re-explained by the experimenter. Once you have answered all the questions correctly, the experiment will start.

Slow evolving societies - Fast evolving leader (SS - FL)

GENERAL INSTRUCTIONS

Hello and welcome to this decision-making experiment!

During this experiment, you can earn money. Your earnings from the experiment will largely depend on your decisions and the decisions of other participants in this experiment. Therefore, it is important to read the instructions carefully.

The amount you will earn at the end of this experiment will be the sum of your earnings during the session and a show-up fee of 5 Euros. If at the end of the session, you have made a loss, you will only receive the show-up fee of 5 Euros. Throughout the experiment, we will not speak of EURO, but rather of ECU (Experimental Currency Units). At the end of the session, the total amount of ECU you earned will be converted into EURO according to the following exchange rate:

$$100 \text{ ECU} = 1 \text{ EURO}$$

All your decisions and answers will be anonymous. You are not allowed to communicate with other participants for the entire duration of the experiment, otherwise, you will be excluded from the session and you will not receive any earnings. You are not allowed to use your phone during the entire session. We kindly ask you to switch it off and put it away.

If you have any questions, please raise your hand or press the red button on the side of your desk. Someone from the experimental team, will come and answer your questions immediately.

The experiment consists of two parts. You will now receive the instructions for Part 1. At the end of Part 1, you will receive the instructions for Part 2.

PART 1

Duration and matching

The first part of the experiment consists of 31 periods. All participants in the session form a group of 21 people. At the beginning of the session, each of you will be randomly assigned a role: Diamond or Club. One person will be assigned the role of Club, while the other 20 people will be assigned the role of Diamonds. You will never change role until the end of the session.

INSTRUCTIONS FOR THE DIAMOND PARTICIPANTS

At the beginning of each period, if you are a Diamond participant, you will be randomly matched to one of the other 19 Diamond participants. You will never know the identity of the person with whom you are matched in a given period.

Your task

Your task in each period is to choose between two actions: **BLUE** or **GREEN**. The participant you are matched with in this period also chooses **BLUE** or **GREEN** at the same time as you. Your earnings from a given period will depend partly on your actions and the action of the other participant with whom you were matched.

There are four possibilities:

- (i) You both choose **BLUE**.
- (ii) You both choose **GREEN**.
- (iii) You choose **BLUE** and the other participant chooses **GREEN**
- (iv) You choose **GREEN** and the other participant chooses **BLUE**

Your Earnings

Your Final Earnings in Part 1 will be the sum of the earnings you made during the 31 periods.

Your earnings in a given period of the experiment depend on two factors: the choices and the types of participants.

There are two possible types in the experiment: **Type A** and **Type B**. We will explain precisely how the types are determined later on. For the time being, it suffices to know that **all participants start off as Type A**.

-A Type A participant receives **30 ECU** when they choose **BLUE** and **20 ECU** when they choose **GREEN**.

-A Type B participant receives **20 ECU** when they choose **BLUE** and **30 ECU** when they choose **GREEN**.

If the other participant chooses the same color as you, the earnings above are the final earnings in a given period. However, every time you and the other participant you were matched with choose **different colors** (i.e., cases iii and iv above) you both receive a “**miscoordination penalty**”. The penalty may differ for the two participants.

In other words, the amount you will receive (see above) will be reduced by **4 ECU for each participant in the group of 20 Diamond participants that chose a different color than you**. That is, the more people choose a different action than you, the greater will be your miscoordination penalty.

You will be informed about the number of participants who chose a different color than you with a delay of one period. Thus, if you and the player who is matched to you chose a different color, you will learn your exact payoff with a delay of one period.

Example 1: Suppose that you and the other participant you were matched with are both Type A. You choose **BLUE** and the other participant chooses **GREEN**. All the other Diamond participants in your group choose **BLUE**, except the participant you are matched with who chose **GREEN**. Then you will earn 30 ECU minus (1 participant * 4 ECU) = 26 ECU. The person choosing **GREEN** will earn 20 ECU minus (19 participants * 4 ECU) = - 56 ECU, that is, this participant will lose 56 ECU.

Example 2: Suppose that you are Type B and choose **GREEN**. Assume further, that the participant you were matched with is Type A and chooses **BLUE**. Among the 20 Diamond participants, you and another participant in your group are the only ones choosing **GREEN**; the rest choose **BLUE**. Then you will earn 30 ECU minus (18 participants * 4 ECU) = -42 ECU. The other participant will earn 30 ECU minus (2 participants * 4 ECU) = 22 ECU.

Example 3: Suppose that you and the other participant are Type B. You choose **BLUE** and the participant you were matched with chooses **GREEN**. Assume further that among all the 20 Diamond participants, 10 choose **GREEN** and 10 choose **BLUE**. Then you will earn 20 ECU minus $(10 \text{ participants} * 4 \text{ ECU}) = -20 \text{ ECU}$. The person choosing **GREEN** will earn 30 ECU minus $(10 \text{ participants} * 4 \text{ ECU}) = -10 \text{ ECU}$.

Type determination and implications for earnings

As mentioned above, **all participants start the experiment as Type A**. The probability of changing type in period 1 is equal to zero. From period 2 onward, each participant has a probability p of becoming Type B. In each period, this probability **increases** with a fixed percentage p .

For 16 Diamond participants, the probability p of becoming Type B is 0.6% in period 2 (i.e., they have 0.6 chances out of 100 of becoming Type B in period 2). This probability **increases** by 0.6 percentage points in each subsequent period. This means that their probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on. For the 4 other Diamond participants, the probability p of becoming Type B is 6% in period 2 (i.e., they have 6 chances in 100 of becoming Type B in period 2). This probability **increases** by 6 percentage points in each subsequent period. This means that their probability of becoming Type B is 12% in period 3, 18% in period 4, and so on.

These probabilities are randomly assigned and are independent for each participant. You will be informed of your probability when the game starts.

The exact period in which you will change Type has already been determined before the beginning of the session. The selection was done independently for each member of your group, based on simulations that reflect a probability of change that increases with each period by: 0.6 percentage points for 16 participants and by 6 percentage points for 4 participants.

If your type changes to Type B you will be informed of this change in the beginning of the period, but you will not know whether the participant with whom you are matched has changed type or is still Type A. The change in type is definitive: once a participant becomes Type B, they do not change back to Type A.

INSTRUCTIONS FOR THE CLUB PARTICIPANT

Your task

At the beginning of each period, if you are a Club participant, you observe your type, the type of each **Diamond** participant in that period, and the number of **Diamond** participants who chose each color in the previous period. Unlike the **Diamond** participants, you do not have to choose a color yourself. Your task is to send a message to the **Diamond** participants. Your message consists of recommending an action, **BLUE** or **GREEN**, to the Diamond participants. Your message will be observed by all **Diamond** participants before they make their choice.

Your Earnings

Your final earnings in Part 1 will be determined based on the earnings obtained in only one of the period of the game. This period will be randomly selected by the program, but you will only be informed of the period selected at the end of the session.

Two factors determine your earnings in a given period of the experiment: your type and the choices made by **Diamond** participants.

You receive:

- **900 ECU** if you are Type A and **all Diamond** participants choose **BLUE**, or if you are Type B and **all Diamond** participants choose **GREEN**
- **400 ECU** if you are Type A and **all Diamond** participants choose **GREEN**, or if you are Type B and **all Diamond** participants choose **BLUE**.

However, every time the **Diamonds participants** do not unanimously choose the same color, you receive a **miscoordination penalty**. This penalty will depend on the situation.

When you are Type A :

- If the majority of **Diamond** participants choose **BLUE**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **GREEN**.
- If the majority of **Diamond** participants choose **GREEN**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **BLUE**.

When you are Type B :

- If the majority of **Diamond** participants choose **GREEN**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **BLUE**.
- If the majority of **Diamond** participants choose **BLUE**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **GREEN**

Example 1: Suppose that you that are Type B. Suppose also that the majority of **Diamond** participants (e.g., 14 people) chose **GREEN**. In this period, you earn $900 - (6 \text{ participants who chose } \text{BLUE} * 70 \text{ UME}) = 480 \text{ UME}$.

Example 2: Suppose that you are Type B. Suppose also that the majority of **Diamond** participants (e.g., 15 people) have chosen **BLUE**. In this period, you earn $400 - (5 \text{ participants who chose } \text{GREEN} * 20 \text{ UME}) = 300 \text{ UME}$.

Type determination and implications for earnings

If you are the Club participant, you begin the experiment as Type A.

In period 2, you have a probability p of 6% of becoming Type B. This probability **increases** by 6 percentage points in each subsequent period. This means that your probability of becoming Type B is 12% in period 3, 18% in period 4, and so on. The exact period in which you change type has already been determined before the start of the session on the basis of simulations that reflect a probability of change that increases in each period by 6 percentage points.

This completes the description of the first part. Please read these instructions again. If you have any questions, raise your hand or press the red button. Otherwise, you can start answering the training questions on the computer. The purpose of these questions is to make sure you have understood the instructions. Any points that are not clear enough will be re-explained by the experimenter. Once you have answered all the questions correctly, the experiment will start.

Fast evolving societies - Slow evolving leader (FS - SL)

GENERAL INSTRUCTIONS

Hello and welcome to this decision-making experiment!

During this experiment, you can earn money. Your earnings from the experiment will largely depend on your decisions and the decisions of other participants in this experiment. Therefore, it is important to read the instructions carefully.

The amount you will earn at the end of this experiment will be the sum of your earnings during the session and a show-up fee of 5 Euros. If at the end of the session, you have made a loss, you will only receive the show-up fee of 5 Euros. Throughout the experiment, we will not speak of EURO, but rather of ECU (Experimental Currency Units). At the end of the session, the total amount of ECU you earned will be converted into EURO according to the following exchange rate:

$$100 \text{ ECU} = 1 \text{ EURO}$$

All your decisions and answers will be anonymous. You are not allowed to communicate with other participants for the entire duration of the experiment, otherwise, you will be excluded from the session and you will not receive any earnings. You are not allowed to use your phone during the entire session. We kindly ask you to switch it off and put it away.

If you have any questions, please raise your hand or press the red button on the side of your desk. Someone from the experimental team, will come and answer your questions immediately.

The experiment consists of two parts. You will now receive the instructions for Part 1. At the end of Part 1, you will receive the instructions for Part 2.

PART 1

Duration and matching

The first part of the experiment consists of 31 periods. All participants in the session form a group of 21 people. At the beginning of the session, each of you will be randomly assigned a role: Diamond or Club. One person will be assigned the role of Club, while the other 20 people will be assigned the role of Diamonds. You will never change role until the end of the session.

INSTRUCTIONS FOR THE DIAMOND PARTICIPANTS

At the beginning of each period, if you are a Diamond participant, you will be randomly matched to one of the other 19 Diamond participants. You will never know the identity of the person with whom you are matched in a given period.

Your task

Your task in each period is to choose between two actions: **BLUE** or **GREEN**. The participant you are matched with in this period also chooses **BLUE** or **GREEN** at the same time as you. Your earnings from a given period will depend partly on your actions and the action of the other participant with whom you were matched.

There are four possibilities:

- (i) You both choose **BLUE**.
- (ii) You both choose **GREEN**.
- (iii) You choose **BLUE** and the other participant chooses **GREEN**
- (iv) You choose **GREEN** and the other participant chooses **BLUE**

Your Earnings

Your Final Earnings in Part 1 will be the sum of the earnings you made during the 31 periods.

Your earnings in a given period of the experiment depend on two factors: the choices and the types of participants.

There are two possible types in the experiment: **Type A** and **Type B**. We will explain precisely how the types are determined later on. For the time being, it suffices to know that **all participants start off as Type A**.

-A Type A participant receives **30 ECU** when they choose **BLUE** and **20 ECU** when they choose **GREEN**.

-A Type B participant receives **20 ECU** when they choose **BLUE** and **30 ECU** when they choose **GREEN**.

If the other participant chooses the same color as you, the earnings above are the final earnings in a given period. However, every time you and the other participant you were matched with choose **different colors** (i.e., cases iii and iv above) you both receive a “**miscoordination penalty**”. The penalty may differ for the two participants.

In other words, the amount you will receive (see above) will be reduced by **4 ECU for each participant in the group of 20 Diamond participants that chose a different color than you**. That is, the more people choose a different action than you, the greater will be your miscoordination penalty.

You will be informed about the number of participants who chose a different color than you with a delay of one period. Thus, if you and the player who is matched to you chose a different color, you will learn your exact payoff with a delay of one period.

Example 1: Suppose that you and the other participant you were matched with are both Type A. You choose **BLUE** and the other participant chooses **GREEN**. All the other Diamond participants in your group choose **BLUE**, except the participant you are matched with who chose **GREEN**. Then you will earn 30 ECU minus (1 participant * 4 ECU) = 26 ECU. The person choosing **GREEN** will earn 20 ECU minus (19 participants * 4 ECU) = - 56 ECU, that is, this participant will lose 56 ECU.

Example 2: Suppose that you are Type B and choose **GREEN**. Assume further, that the participant you were matched with is Type A and chooses **BLUE**. Among the 20 Diamond participants, you and another participant in your group are the only ones choosing **GREEN**; the rest choose **BLUE**. Then you will earn 30 ECU minus (18 participants * 4 ECU) = -42 ECU. The other participant will earn 30 ECU minus (2 participants * 4 ECU) = 22 ECU.

Example 3: Suppose that you and the other participant are Type B. You choose **BLUE** and the participant you were matched with chooses **GREEN**. Assume further that among all the 20 Diamond participants, 10 choose **GREEN** and 10 choose **BLUE**. Then you will earn 20 ECU minus (10 participants * 4 ECU) = -20 ECU. The person choosing **GREEN** will earn 30 ECU minus (10 participants * 4 ECU) = -10 ECU.

Type determination and implications for earnings

As mentioned above, **all participants start the experiment as Type A**. The probability of changing type in period 1 is equal to zero. From period 2 onward, each participant has a probability p of becoming Type B. In each period, this probability **increases** with a fixed percentage p .

For 16 Diamond participants, the probability p of becoming Type B is 6% in period 2 (i.e., they have 6 chances out of 100 of becoming Type B in period 2). This probability **increases** by 6 percentage points in each subsequent period. This means that their probability of becoming Type B is 12% in period 3, 18% in period 4, and so on. For the 4 other Diamond participants, the probability p of becoming Type B is 0.6% in period 2 (i.e., they have 0.6 chances in 100 of becoming Type B in period 2). This probability **increases** by 0.6 percentage points in each subsequent period. This means that their probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on.]

These probabilities are randomly assigned and are independent for each participant. You will be informed of your probability when the game starts.

The exact period in which you will change Type has already been determined before the beginning of the session. The selection was done independently for each member of your group, based on simulations that reflect a probability of change that increases with each period by: 6 percentage points for 16 participants and by 0.6 percentage points for 4 participants.

If your type changes to Type B you will be informed of this change in the beginning of the period, but you will not know whether the participant with whom you are matched has changed type or is still Type A. The change in type is definitive: once a participant becomes Type B, they do not change back to Type A.

INSTRUCTIONS FOR THE CLUB PARTICIPANT

Your task

At the beginning of each period, if you are a Club participant, you observe your type, the type of each **Diamond** participant in that period, and the number of **Diamond** participants who chose each color in the previous period. Unlike the **Diamond** participants, you do not have to choose a color yourself. Your task is to send a message to the **Diamond** participants. Your message consists of recommending an action, **BLUE** or **GREEN**, to the Diamond participants. Your message will be observed by all **Diamond** participants before they make their choice.

Your Earnings

Your final earnings in Part 1 will be determined based on the earnings obtained in only one of the period of the game. This period will be randomly selected by the program, but you will only be informed of the period selected at the end of the session.

Two factors determine your earnings in a given period of the experiment: your type and the choices made by **Diamond** participants.

You receive:

- **900 ECU** if you are Type A and **all Diamond** participants choose **BLUE**, or if you are Type B and **all Diamond** participants choose **GREEN**
- **400 ECU** if you are Type A and **all Diamond** participants choose **GREEN**, or if you are Type B and **all Diamond** participants choose **BLUE**.

However, every time the **Diamonds participants** do not unanimously choose the same color, you receive a **miscoordination penalty**. This penalty will depend on the situation.

When you are Type A :

- If the majority of **Diamond** participants choose **BLUE**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **GREEN**.
- If the majority of **Diamond** participants choose **GREEN**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **BLUE**.

When you are Type B :

- If the majority of **Diamond** participants choose **GREEN**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **BLUE**.
- If the majority of **Diamond** participants choose **BLUE**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **GREEN**

Example 1: Suppose that you that are Type B. Suppose also that the majority of **Diamond** participants (e.g., 14 people) chose **GREEN**. In this period, you earn $900 - (6 \text{ participants who chose } \text{BLUE} * 70 \text{ UME}) = 480 \text{ UME}$.

Example 2: Suppose that you are Type B. Suppose also that the majority of **Diamond** participants (e.g., 15 people) have chosen **BLUE**. In this period, you earn $400 - (5 \text{ participants who chose } \text{GREEN} * 20 \text{ UME}) = 300 \text{ UME}$.

Type determination and implications for earnings

If you are the Club participant, you begin the experiment as Type A.

In period 2, you have a probability p of 0.6% of becoming Type B. This probability **increases** by 0.6 percentage points in each subsequent period. This means that your probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on. The exact period in which you change type has already been determined before the start of the session based on simulations that reflect a probability of change that increases in each period by 0.6 percentage points.

This completes the description of the first part. Please read these instructions again. If you have any questions, raise your hand or press the red button. Otherwise, you can start answering the training questions on the computer. The purpose of these questions is to make sure you have understood the instructions. Any points that are not clear enough will be re-explained by the experimenter. Once you have answered all the questions correctly, the experiment will start.

Fast evolving Societies - Slow evolving Leader - Communication (FS-SL-COMM)

GENERAL INSTRUCTIONS

Hello and welcome to this decision-making experiment!

During this experiment, you can earn money. Your earnings from the experiment will largely depend on your decisions and the decisions of other participants in this experiment. Therefore, it is important to read the instructions carefully.

The amount you will earn at the end of this experiment will be the sum of your earnings during the session and a show-up fee of 5 Euros. If at the end of the session, you have made a loss, you will only receive the show-up fee of 5 Euros. Throughout the experiment, we will not speak of EURO, but rather of ECU (Experimental Currency Units). At the end of the session, the total amount of ECU you earned will be converted into EURO according to the following exchange rate:

$$100 \text{ ECU} = 1 \text{ EURO}$$

All your decisions and answers will be anonymous. You are not allowed to communicate with other participants for the entire duration of the experiment, otherwise, you will be excluded from the session and you will not receive any earnings. You are not allowed to use your phone during the entire session. We kindly ask you to switch it off and put it away.

If you have any questions, please raise your hand or press the red button on the side of your desk. Someone from the experimental team, will come and answer your questions immediately.

The experiment consists of two parts. You will now receive the instructions for Part 1. At the end of Part 1, you will receive the instructions for Part 2.

PART 1

Duration and matching

The first part of the experiment consists of 31 periods. All participants in the session form a group of 21 people. At the beginning of the session, each of you will be randomly assigned a role: Diamond or Club. One person will be assigned the role of Club, while the other 20 people will be assigned the role of Diamonds. You will never change role until the end of the session.

INSTRUCTIONS FOR THE DIAMOND PARTICIPANTS

At the beginning of each period, if you are a Diamond participant, you will be randomly matched to one of the other 19 Diamond participants. You will never know the identity of the person with whom you are matched in a given period.

Your task

Your task in each period is to choose between two actions: **BLUE** or **GREEN**. The participant you are matched with in this period also chooses **BLUE** or **GREEN** at the same time as you. Your earnings from a given period will depend partly on your actions and the actions of the other participant with whom you were matched.

There are four possibilities:

- (i) You both choose **BLUE**.
- (ii) You both choose **GREEN**.
- (iii) You choose **BLUE** and the other participant chooses **GREEN**
- (iv) You choose **GREEN** and the other participant chooses **BLUE**

Your Earnings

Your Final Earnings in Part 1 will be the sum of the earnings you made during the 31 periods.

Your earnings in a given period of the experiment depend on two factors: the choices and the types of participants.

There are two possible types in the experiment: **Type A** and **Type B**. We will explain precisely how the types are determined later on. For the time being, it suffices to know that **all participants start off as Type A**.

-A Type A participant receives **30 ECU** when they choose **BLUE** and **20 ECU** when they choose **GREEN**.

-A Type B participant receives **20 ECU** when they choose **BLUE** and **30 ECU** when they choose **GREEN**.

If the other participant chooses the same color as you, the earnings above are the final earnings in a given period. However, every time you and the other participant you were matched with choose **different colors** (i.e., cases iii and iv above) you both receive a “**miscoordination penalty**”. The penalty may differ for the two participants.

In other words, the amount you will receive (see above) will be reduced by **4 ECU for each participant in the group of 20 Diamond participants that chose a different color than you**. That is, the more people choose a different action than you, the greater will be your miscoordination penalty.

At the beginning of periods 5, 10, 15, 20, 25, and 30, if you are a Diamond participant, you will be asked to recommend an action to your fellow Diamond participants: **BLUE** or **GREEN**. Then, the program will randomly draw 10 recommendations from all the recommendations made by all the Diamond participants. These 10 recommendations will be shown to all the Diamonds participants, including yourself before you make your decision. The Club participant will never be informed of the content of these recommendations.

You will be informed about the number of participants who chose a different color than you with a delay of one period. Thus, if you and the player who is matched to you chose a different color, you will learn your exact payoff with a delay of one period.

Example 1: Suppose that you and the other participant you were matched with are both Type A. You choose **BLUE** and the other participant chooses **GREEN**. All the other Diamond participants in your group choose **BLUE**, except the participant you are matched with who chose **GREEN**. Then you will earn 30 ECU minus (1 participant * 4 ECU) = 26 ECU. The person choosing **GREEN** will earn 20 ECU minus (19 participants * 4 ECU) = - 56 ECU, that is, this participant will lose 56 ECU.

Example 2: Suppose that you are Type B and choose **GREEN**. Assume further, that the participant you were matched with is Type A and chooses **BLUE**. Among the 20 Diamond participants, you and another participant in your group are the only ones choosing **GREEN**; the rest choose **BLUE**. Then you will earn 30 ECU minus (18 participants * 4 ECU) = -42 ECU. The other participant will earn 30 ECU minus (2 participants * 4 ECU) = 22 ECU.

Example 3: Suppose that you and the other participant are Type B. You choose **BLUE** and the participant you were matched with chooses **GREEN**. Assume further that among all the 20 Diamond participants, 10 choose **GREEN** and 10 choose **BLUE**. Then you will earn 20 ECU minus (10 participants * 4 ECU) = -20 ECU. The person choosing **GREEN** will earn 30 ECU minus (10 participants * 4 ECU) = - 10 ECU.

Type determination and implications for earnings

As mentioned above, **all participants start the experiment as Type A**. The probability of changing type in period 1 is equal to zero. From period 2 onward, each participant has a probability p of becoming Type B. In each period, this probability **increases** with a fixed percentage p .

For 16 Diamond participants, the probability p of becoming Type B is 6% in period 2 (i.e., they have 6 chances out of 100 of becoming Type B in period 2). This probability **increases** by 6 percentage points in each subsequent period. This means that their probability of becoming Type B is 12% in period 3, 18% in period 4, and so on. For the 4 other Diamond participants, the probability p of becoming Type B is 0.6% in period 2 (i.e., they have 0.6 chances in 100 of becoming Type B in period 2). This probability **increases** by 0.6 percentage points in each subsequent period. This means that their probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on.

These probabilities are randomly assigned and are independent for each participant. You will be informed of your probability when the game starts.

The exact period in which you will change Type has already been determined before the beginning of the session. The selection was done independently for each member of your group, based on simulations that reflect a probability of change that increases with each period by: 6 percentage points for 16 participants and by 0.6 percentage points for 4 participants.

If your type changes to Type B you will be informed of this change at the beginning of the period, but you will not know whether the participant with whom you are matched has changed type or is still Type A. The change in type is definitive: once a participant becomes Type B, they do not change back to Type A.

INSTRUCTIONS FOR THE CLUB PARTICIPANT

Your task

At the beginning of each period, if you are a Club participant, you observe your type, the type of each **Diamond** participant in that period, and the number of **Diamond** participants who chose each color in the previous period. Unlike the **Diamond** participants, you do not have to choose a color yourself. Your task is to send a message to the **Diamond** participants. Your message consists of recommending an action, **BLUE** or **GREEN**, to the **Diamond** participants. Your message will be observed by all **Diamond** participants before they make their choice.

Your Earnings

Your final earnings in Part 1 will be determined based on the earnings obtained in only one of the periods of the game. This period will be randomly selected by the program, but you will only be informed of the period selected at the end of the session.

Two factors determine your earnings in a given period of the experiment: your type and the choices made by **Diamond** participants.

You receive:

- **900 ECU** if you are Type A and **all Diamond** participants choose **BLUE**, or if you are Type B and **all Diamond** participants choose **GREEN**
- **400 ECU** if you are Type A and **all Diamond** participants choose **GREEN**, or if you are Type B and **all Diamond** participants choose **BLUE**.

However, every time the **Diamonds participants** do not unanimously choose the same color, you receive a **miscoordination penalty**. This penalty will depend on the situation.

When you are Type A :

- If the majority of **Diamond** participants choose **BLUE**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **GREEN**.
- If the majority of **Diamond** participants choose **GREEN**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **BLUE**.

When you are Type B :

- If the majority of **Diamond** participants choose **GREEN**, you earn 900 ECU minus a penalty of 70 ECU for each **Diamond** participant who chooses **BLUE**.
- If the majority of **Diamond** participants choose **BLUE**, you earn 400 ECU minus a penalty of 20 ECU for each **Diamond** participant who chooses **GREEN**.

Example 1: Suppose that you are Type B. Suppose also that the majority of **Diamond** participants (e.g., 14 people) chose **GREEN**. In this period, you earn $900 - (6 \text{ participants who chose BLUE} * 70 \text{ UME}) = 480 \text{ UME}$.

Example 2: Suppose that you are Type B. Suppose also that the majority of **Diamond** participants (e.g., 15 people) have chosen **BLUE**. In this period, you earn $400 - (5 \text{ participants who chose GREEN} * 20 \text{ UME}) = 300 \text{ UME}$.

Type determination and implications for earnings

If you are the Club participant, you begin the experiment as Type A.

In period 2, you have a probability p of 0.6% of becoming Type B. This probability **increases** by 0.6 percentage points in each subsequent period. This means that your probability of becoming Type B is 1.2% in period 3, 1.8% in period 4, and so on. The exact period in which you change type has already been determined before the start of the session based on simulations that reflect a probability of change that increases in each period by 0.6 percentage points.

This completes the description of the first part. Please read these instructions again. If you have any questions, raise your hand or press the red button. Otherwise, you can start answering the training questions on the computer. The purpose of these questions is to make sure you have understood the instructions. Any points that are not clear enough will be re-explained by the experimenter. Once you have answered all the questions correctly, the experiment will start.

PART 2 [*Displayed on computer screens after completion of Part 1 - All treatments*]

In Part 2 of this experiment, you will decide how to allocate 50 balls between two buckets. Your task is to put each ball, one by one, into one of the two buckets: the yellow bucket or the violet bucket.

The rule is to put the balls in the violet bucket.

The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice.

For each ball you put in the violet bucket, you will receive 1 ECU, and for each ball you put in the yellow bucket, you will receive 2 ECU.

Once the experiment starts, you will have **5** minutes to put the balls into the buckets. When you are finished, please wait quietly until the end of the 5-minute period.

Your payment from this part will be the sum of payments from the yellow and violet buckets.

If you have any questions, please raise your hand and an experimenter will answer them privately. Otherwise, please click OK to continue.

QUESTIONNAIRES *[Displayed on computer screens after completion of Part 2 - All treatments]*

Several questionnaires will appear on your screen. Please answer them truthfully. We remind you that your answers are confidential and anonymous. Click OK to display the questions.

Please indicate your opinion for each of the following statements concerning your attitudes in life in general.

1. Are you generally an impatient person, or someone who always shows great patience? Please indicate your response on a 1 to 10 scale, where the value 1 means: “very impatient” and the value 10 means “very patient”.
2. To what extent do you trust people you have just met? Please indicate your response on a 1 to 10 scale, where the value 1 means: “not at all” and the value 10 means “completely”.
3. Are you generally willing to take risks, or do you try to avoid risks? Please indicate your response on a 1 to 10 scale, where the value 1 means: “not willing to take risks at all” and the value 10 means “fully ready to take risks”.
4. Are your social and economic ideas more liberal or conservative? Indicate your answer on a scale of 1 to 10, with 1 for “very conservative” and 10 for “very liberal”.
5. You will be presented with 7 adjective pairs. For each pair, select one of the 5 buttons. The buttons from left to right mean that the adjective on the left-hand side describes you (1) better, (2) slightly better, (3) equally well, (4) slightly worse, (5) worse than the adjective on the right-hand side. That is, the closer your choice to an adjective, the better you think that adjective describes you as a person.

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence.

- compliant vs. defiant
 - resistant vs. acquiescent
 - inflexible vs. adaptive
 - agreeing vs. disagreeing
 - cooperative vs. uncooperative
 - opposing vs. accommodating
 - differing vs. concurring
6. What year were you born?
 7. What is your gender:
 - Male
 - Female
 - Neutral

- Student
- Employee
- Self-employed
- Unemployed
- Retired/Inactive

9. (*Gate-Lab sessions*: If you are a student, what university or school are you currently going to?)

- Ecole Centrale de Lyon
- emlyon business school
- ISOSTEO
- ITECH
- Faculty of Economics
- Other

OR

10. (*GAEL sessions*: If you are a student, what university or school are you currently going to?)

- Engineering School
- Management or Business School
- Medicine
- IUT (Technology)
- Faculty of Economics
- Other

11. According to you, what does this experiment intend to test?

12. If you could leave an advice to another participant in your role, what strategy would you recommend?

13. Now that the experiment is over, starting from which period do you think you should have started to choose Green?

References

1. J Andreoni, N Nikiforakis, S Siegenthaler, Predicting social tipping and norm change in controlled experiments. *Proc. Natl. Acad. Sci.* **118** (2021).